GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 16, 2004, 23:30:45 ; Search time 529 Seconds
(without alignments)
2589.979 Million cell updates/sec

Title: US-10-043-649-2
Perfect score: 1351
Sequence: 1 MGSLPSRRKSLPSPSLSSSV.....RESLSFYISLNDEAVSLDDA 261

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10043649/runat_16112004_060535_28709/app_query.fasta_1.455
-DB=N_Geneseq_23Sep04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10043649_@CGN_1_1_470_@runat_16112004_060535_28709 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_23Sep04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1351	100.0	786	6	AAL44089	Aal44089 Human mod
2	1351	100.0	786	6	ABQ74343	Abq74343 Human Src
3	1351	100.0	2567	6	AAD43980	Aad43980 Human Src
c 4	1351	100.0	3646	10	ADF90741	Adf90741 Human hep
5	1347	99.7	1183	6	ABK61465	Abk61465 Human cDN
6	1345	99.6	2788	10	ADF82458	Adf82458 Leukaemia
7	1273	94.2	837	3	AAC77202	Aac77202 Human ORF
8	1210.5	89.6	1413	6	ABQ99374	Abq99374 Human cod
9	1200.5	88.9	737	6	AAL44090	Aal44090 Mouse MAR
10	1132.5	83.8	2049	5	AAS74750	Aas74750 DNA encod
11	1032	76.4	1348	6	AAL44087	Aal44087 Mouse mod
12	830	61.4	763	6	ABQ98670	Abq98670 Human ORF
13	731	54.1	603	5	AAS74748	Aas74748 DNA encod
14	647	47.9	864	5	ADL63090	Adl63090 Human ova
15	586	43.4	875	6	ABQ99151	Abq99151 Human ORF
16	492	36.4	3756	8	ABX62975	Abx62975 Human act
17	491	36.3	2109	4	AAS02049	Aas02049 DNA encod
18	491	36.3	2665	6	ABL65189	Abl65189 Lung canc
19	491	36.3	2665	6	ABK83738	Abk83738 Human cDN
20	491	36.3	2665	8	ACC81091	Acc81091 Human Src
21	491	36.3	2665	12	ADL83115	Adl83115 Human PRO
22	491	36.3	2665	12	ADP12833	Adp12833 Reference
23	491	36.3	3452	12	ADQ22520	Adq22520 Human sof
24	452.5	33.5	444	6	ABQ98669	Abq98669 Human ORF
25	374.5	27.7	2298	6	ABK83935	Abk83935 Human cDN
26	374.5	27.7	2298	10	ADD19015	Add19015 Human dis
27	374.5	27.7	2298	10	ADF81586	Adf81586 Leukaemia
28	374.5	27.7	2298	11	ADI31832	Adi31832 Human cDN
29	374.5	27.7	2298	11	ADN95431	Adn95431 Human BEC
30	374.5	27.7	2298	12	ADL22891	Adl22891 Human MP2
31	374.5	27.7	2298	12	ADN04497	Adn04497 Antipsori
32	374.5	27.7	2298	12	ADP10425	Adp10425 Reference
33	374.5	27.7	4175	10	ADF81585	Adf81585 Leukaemia
34	364.5	27.0	1924	12	ADI04091	Adi04091 Human HCK
35	364.5	27.0	1926	6	ABK83940	Abk83940 Human cDN
36	364.5	27.0	1926	12	ADJ71658	Adj71658 Human NOV
37	364.5	27.0	2015	6	ABL66673	Abl66673 Lung canc
38	364.5	27.0	2015	6	ABK83939	Abk83939 Human cDN
39	364.5	27.0	2015	10	AAD62155	Aad62155 Human hae
40	364.5	27.0	2015	11	ADI31779	Adi31779 Human cDN
41	364.5	27.0	2015	12	ADL22888	Adl22888 Human MP2
42	364.5	27.0	2015	12	ADP12819	Adp12819 Reference
43	364.5	27.0	2025	11	ADL71038	Adl71038 Gene enco
44	364.5	27.0	2341	10	ADE53793	Ade53793 Human pro
45	362.5	26.8	1911	6	ABK63704	Abk63704 Rat seque

ALIGNMENTS

RESULT 1
AAL44089
ID AAL44089 standard; cDNA; 786 BP.
XX
AC AAL44089;
XX
DT 03-OCT-2002 (first entry)
XX
DE Human modulator of antigen receptor signalling protein coding sequence.
XX
KW Human; gene; ss; gene therapy; modulator of antigen receptor signalling;
KW MARS; tumour suppressor gene; Scr-like adaptor protein; SLAP;
KW myeloid malignancy; acute myelogenous leukaemia; autoimmune disorder;
KW immunosuppression; myeloproliferative disorder; breast cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1. .786
FT /*tag= a
FT /product= "Human MARS protein"
XX
PN WO200242452-A2.
XX
PD 30-MAY-2002.
XX
PF 26-NOV-2001; 2001WO-CA001662.
XX
PR 27-NOV-2000; 2000CA-02324663.
XX

PA (HOSP-) HOSPITAL FOR SICK CHILDREN.
XX
PI Mcglade JC, Loreto MP;
XX
DR WPI; 2002-566564/60.
DR P-PSDB; AAO15457.
XX
PT New isolated modulator of antigen receptor signaling protein or its
PT fragment, useful for treating malignant disorders such as myeloid
PT malignancies, autoimmune disorders and myeloproliferative disorders.
XX
PS Claim 12; Page 75; 110pp; English.
XX
CC The invention comprises the amino acid and coding sequences of modulator
CC of antigen receptor signalling (MARS) proteins. The MARS protein is a
CC putative tumour suppressor gene and exhibits structural and sequence
CC similarity to the Scr-like adaptor protein (SLAP). The MARS DNA and
CC protein sequences of the invention are useful for the treatment of
CC myeloid malignancies (e.g. acute myelogenous leukaemia) autoimmune
CC disorders, immunosuppression, myeloproliferative disorders and
CC malignancies related to the de-regulation of tyrosine kinases (e.g.
CC breast cancer). The present cDNA sequence encodes a human MARS protein
XX
SQ Sequence 786 BP; 162 A; 234 C; 231 G; 159 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.36e-110	Length:	786
Score:	1351.00	Matches:	261
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-043-649-2 (1-261) x AAL44089 (1-786)

```
Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660

Qy        221 AspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGly 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720

Qy        241 LeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAspAsp 260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780

Qy        261 Ala 261
              |||
Db        781 GCC 783
```

RESULT 2
ABQ74343
ID ABQ74343 standard; cDNA; 786 BP.
XX
AC ABQ74343;
XX
DT 15-OCT-2002 (first entry)
XX
DE Human Src-like inhibitory molecule (SLIM) encoding cDNA.
XX
KW Human; Src-like inhibitory molecule; SLIM; Src-like adapter protein;
KW SLAP; inhibitor; antiinflammatory; immunosuppressive; anti-HIV;
KW modulator; lymphocyte; Cbl; gene therapy; immunodeficiency disorder;
KW acquired immune deficiency syndrome; AIDS; acute inflammatory disorder;
KW chronic inflammatory disorder; autoimmune disorder; transplant rejection;
KW gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1. .786
FT /*tag= a
FT /product= "SLIM"
FT /note= "Src-like inhibitory molecule"
XX
PN WO200255707-A2.
XX
PD 18-JUL-2002.
XX
PF 10-JAN-2002; 2002WO-US000718.
XX
PR 10-JAN-2001; 2001US-0260953P.
XX
PA (RIGE-) RIGEL PHARM INC.
XX
PI Holland SJ, Mendenhall MK, Pardo J, Spencer C, Fu AC, Luo Y;
PI Payan DG, Mancebo HS, Wu J, Zhou X, Shen M, Liao XC, Sheng N;
XX
DR WPI; 2002-575432/61.
DR P-PSDB; ABP52187.
XX
PT New src-like inhibitory molecule protein, useful for treating
PT immunodeficiency disorders and inflammatory disorders, comprises N-
PT terminal myristylation sequence, SH2 domain and/or SH3 domain.
XX
PS Claim 3; Fig 2A; 91pp; English.
XX
CC The present sequence encodes the human Src-like inhibitory molecule
CC (SLIM) protein (I). The present invention describes a SLIM protein
CC comprising an N-terminal myristylation sequence, an N-terminal SH2
CC domain, and an N-terminal SH3 domain which can bind to Cbl, or comprising
CC an N-terminal myristylation sequence and an N-terminal SH2 domain which
CC is unable to bind to Cbl. (I) has antiinflammatory, immunosuppressive and
CC anti-HIV activities, and can be used as a modulator of lymphocyte
CC activation, and of ubiquitination of a Cbl target protein, and in gene

```
CC   therapy. (I) is useful for screening a bioactive agent capable of binding
CC   to SLIM. (I) is also useful for screening a bioactive agent capable of
CC   modulating SLIM binding. (I) or its fragments is useful in the study or
CC   in the treatment of conditions which involves this function or
CC   dysregulation of SLIM protein activity, i.e. to diagnose, treat or
CC   prevent SLIM associated disorders. (I) or the polynucleotide encoding it
CC   (II) is useful for modulating leukocyte and/or platelet activation, for
CC   modulating antigen receptor-induced signalling and activation in
CC   leukocyte and/or platelets and for modulating antigen receptor-induced
CC   signalling and activation in lymphocytes and/or mast cells. (I) or (II)
CC   is also useful for modulating the basal activity of lymphocytes. (I) or
CC   (II) is useful in the treatment of immunodeficiency disorders, such as
CC   acquired immunodeficiency syndrome (AIDS), for the prevention and
CC   treatment of acute inflammatory disorders, chronic inflammatory
CC   disorders, autoimmune disorder and transplant rejection
XX
SQ   Sequence 786 BP; 162 A; 234 C; 231 G; 159 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                1.36e-110      Length:          786
Score:                    1351.00        Matches:         261
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       6              Gaps:            0

US-10-043-649-2 (1-261) x ABQ74343 (1-786)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660

Qy        221 AspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGly 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720

Qy        241 LeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAspAsp 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780

Qy        261 Ala 261
              |||
Db        781 GCC 783

RESULT 3
AAD43980
ID   AAD43980 standard; cDNA; 2567 BP.
XX
AC   AAD43980;
XX
DT   13-DEC-2002  (first entry)
XX
DE   Human Src-Like Adapter Protein-2 (hSLAP-2) cDNA.
XX
KW   Human; SH2/SH3-domain-containing adapter; Src-Like Adapter Protein-2;
KW   SLAP-2; immune disorder; signal transduction; autoimmune disease; cancer;
KW   neoplasm; inflammation disorder; rheumatoid arthritis; osteoarthritis;
KW   psoriasis; rhinitis; allergy; inflammatory bowel disease; gene therapy;
KW   Crohn's disease; systemic lupus erythematosus; tissue/organ rejection;
KW   multiple sclerosis; asthma; acute respiratory distress syndrome;
KW   pulmonary disorder; dermatological; neuroprotective; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             415. .1200
FT                   /*tag=  a
FT                   /product= "Human SLAP-2"
XX
PN   WO200242457-A1.
XX
PD   30-MAY-2002.
XX
PF   20-NOV-2001; 2001WO-US043367.
XX
PR   22-NOV-2000; 2000US-0252545P.
XX
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI   Chang H,  Yang W,  Wu Y,  Whitney GS,  Perez-Villar JJ,  Kanner SB;
XX
DR   WPI; 2002-463632/49.
DR   P-PSDB; AAE26357.
XX
PT   Novel substantially purified human SH2/SH3-domain-containing adapter
PT   polypeptide, termed Src-Like Adapter Protein-2, useful for therapeutic
PT   intervention in immunological and inflammatory disorders and cancer.
XX
PS   Claim 2; Fig 1; 85pp; English.
XX
CC   The invention relates to a substantially purified human SH2/SH3-domain-
CC   containing adapter polypeptide, termed Src-Like Adapter Protein-2 (SLAP-
CC   2). The invention is useful for treating an immune disorder involving
CC   hyperactivity of B- or T- lymphocytes in a mammal. The invention is
CC   useful for screening for antagonists or inhibitors of the interaction of
CC   hSLAP-2 with cellular signalling compounds, for diagnosing, treating or
CC   preventing diseases or disorders associated with aberrant or uncontrolled
CC   cellular signal transduction, for determining those cellular signalling
CC   molecules which associate with hSLAP-2 and which provide critical signals
CC   for cell activation, and as effectors in methods to affect T- cell
CC   activation. The invention is useful in screening assays to identity and
CC   detect candidate bioactive agents that modulate hSLAP-2 bioactivity, for
CC   potential use to treat autoimmune diseases which may be caused by
CC   hyperactivated B cells, as well as to treat diseases which may be caused
CC   by hyperactivated T cells, in addition to other immune system related
CC   conditions, diseases, or disorders, T-cell and B-cell neoplasms,
```

```
CC   inflammation disorders, diseases and conditions, rheumatoid arthritis,
CC   osteoarthritis, psoriasis, rhinitis, inflammatory bowel disease (Crohn's
CC   and ulcerative colitis), allergies, particularly those involving
CC   hyperactivity of B-cells and T- cells, or other immune cells, such as
CC   mast cells or eosinophils, autoimmune diseases such as systemic lupus
CC   erythematosus and multiple sclerosis, pulmonary diseases including
CC   asthma, acute respiratory distress syndrome, and chronic obstructive
CC   pulmonary disorder, tissue/ organ rejection and cancer. The invention is
CC   useful in gene therapy. The present sequence is human SLAP-2 cDNA
XX
SQ   Sequence 2567 BP; 611 A; 741 C; 666 G; 549 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              6.29e-110      Length:          2567
Score:                  1351.00        Matches:         261
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            100.00%        Indels:          0
DB:                     6              Gaps:            0

US-10-043-649-2 (1-261) x AAD43980 (1-2567)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        415 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 474

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        475 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 534

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        535 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 594

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        595 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 654

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        655 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 714

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        715 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 774

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        775 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 834

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        835 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 894

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        895 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 954

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        955 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 1014

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1015 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 1074

Qy        221 AspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGly 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1075 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 1134

Qy        241 LeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAspAsp 260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1135 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 1194
```

```
Qy         261 Ala 261
                |||
Db        1195 GCC 1197

RESULT 4
ADF90741/c
ID   ADF90741 standard; DNA; 3646 BP.
XX
AC   ADF90741;
XX
DT   26-FEB-2004  (first entry)
XX
DE   Human hepatic-fibrosis disease marker SEQ ID 203.
XX
KW   Hepatic fibrosis; marker; chronic hepatitis; liver cirrhosis;
KW   hepatic carcinoma; human; ds.
XX
OS   Homo sapiens.
XX
PN   JP2003259877-A.
XX
PD   16-SEP-2003.
XX
PF   11-MAR-2002; 2002JP-00065013.
XX
PR   11-MAR-2002; 2002JP-00065013.
XX
PA   (SUMU ) SUMITOMO SEIYAKU KK.
XX
DR   WPI; 2003-821598/77.
XX
PT   Hepatic fibrosis disease markers comprising polynucleotides or
PT   antibodies, useful for improved diagnosis, screening and developing drugs
PT   to treat hepatitis, to control cirrhosis and carcinoma.
XX
PS   Claim 1; SEQ ID NO 203; 313pp; Japanese.
XX
CC   The present invention relates to hepatic-fibrosis disease markers
CC   (ADF90539-ADF90871) and related proteins (ADF90872-ADF90917). The
CC   sequences are useful for detecting and treating hepatic fibrosis caused
CC   by alcohol consumption, virus infection, etc., and the associated chronic
CC   hepatitis, etc. leading to liver cirrhosis and hepatic carcinoma. The
CC   markers allow the cause of hepatic fibrosis to be clarified (diagnostic
CC   precision), so more suitable treatments can be developed and given.
XX
SQ   Sequence 3646 BP; 782 A; 954 C; 1031 G; 877 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.:              9.91e-110      Length:          3646
Score:                  1351.00        Matches:         261
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            100.00%        Indels:          0
DB:                     10             Gaps:            0

US-10-043-649-2 (1-261) x ADF90741 (1-3646)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3202 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 3143

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3142 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 3083

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3082 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 3023

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3022 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 2963
```

```
Qy        81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2962 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 2903

Qy       101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2902 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 2843

Qy       121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2842 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 2783

Qy       141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2782 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 2723

Qy       161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2722 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 2663

Qy       181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2662 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 2603

Qy       201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2602 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 2543

Qy       221 AspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGly 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2542 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 2483

Qy       241 LeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAspAsp 260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2482 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 2423

Qy       261 Ala 261
             |||
Db      2422 GCC 2420
```

RESULT 5
ABK61465
ID ABK61465 standard; cDNA; 1183 BP.
XX
AC ABK61465;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human cDNA encoding protein NOV13.
XX
KW Human; gene; ss; NOVX; gene therapy; cardiomyopathy; atherosclerosis;
KW cell signal processing disorder; metabolic pathway modulation disorder;
KW diabetes; cancer; adenocarcinoma; lymphoma; prostate cancer;
KW uterus cancer; immune response; graft-versus-host disease;
KW acquired immunodeficiency syndrome; AIDS; asthma; Crohn's disease;
KW hypertension; congenital heart defects; multiple sclerosis; inflammation;
KW Albright hereditary osteodystrophy.
XX
OS Homo sapiens.
XX
PN WO200216599-A2.
XX
PD 28-FEB-2002.
XX
PF 27-AUG-2001; 2001WO-US026510.
XX
PR 25-AUG-2000; 2000US-0228191P.
PR 08-FEB-2001; 2001US-0267300P.
PR 20-FEB-2001; 2001US-0269961P.
PR 20-MAR-2001; 2001US-0277337P.
XX
PA (CURA-) CURAGEN CORP.
PA (CORT-) COR THERAPEUTICS INC.
XX
PI Burgess CE, Conley PB, Grosse WM, Hart M, Kekuda R, Shimkets RA;
PI Spytek KA, Szekeres ES, Tomlinson JE, Topper JN, Yang R;
XX
DR WPI; 2002-280937/32.
DR P-PSDB; AAU91308.
XX
PT New polypeptides for treating or preventing a disorder associated with
PT them, in humans, e.g. cardiomyopathy, atherosclerosis or cancers.
XX
PS Claim 1; Page 98; 263pp; English.
XX
CC The invention relates to an isolated polypeptide (NOVX) a mature form of
CC NOVX, a NOVX variant (differing by no more than 15%), the nucleotide
CC encoding NOVX (or its complement, fragment or variant). NOVX is NOV1-14,
CC 15a, 15b, 16a, and 16b. The NOVX polypeptide, nucleic acid encoding it
CC and antibody against it, are useful for treating or preventing (e.g. by
CC gene therapy) a NOVX-associated disorder in humans, e.g. cardiomyopathy,
CC atherosclerosis, a disorder related to cell signal processing and
CC metabolic pathway modulation, diabetes or cancers. The NOVX polypeptide
CC and nucleic acids are also useful for determining the presence of
CC predisposition to the diseases. The NOVX nucleic acid and polypeptide are
CC especially useful in therapeutic or prophylactic applications for
CC disorders associated with aberrant NOVX expression or activity, e.g.
CC cancers (e.g. adenocarcinoma, lymphoma, prostate cancer or uterus
CC cancer), immune response, graft-versus-host disease, acquired
CC immunodeficiency syndrome (AIDS), asthma, Crohn's disease, hypertension,
CC congenital heart defects, multiple sclerosis, inflammation or Albright
CC hereditary osteodystrophy and many other diseases listed in the
CC specification. The DNA encoding the protein is useful in gene therapy for
CC treating the conditions. This is also useful in detection assays,
CC chromosome mapping, tissue typing, diagnostic or prognostic assays, or
CC for developing a powerful assay system for functional analysis of various
CC human disorders, as well as in diagnostic applications. The present
CC sequence encodes a NOVX protein
XX
SQ Sequence 1183 BP; 251 A; 359 C; 333 G; 240 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5.22e-110	Length:	1183
Score:	1347.00	Matches:	260
Percent Similarity:	99.62%	Conservative:	0
Best Local Similarity:	99.62%	Mismatches:	1
Query Match:	99.70%	Indels:	0
DB:	6	Gaps:	0

US-10-043-649-2 (1-261) x ABK61465 (1-1183)

```
Qy         1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerServal 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       398 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 457

Qy        21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       458 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 517

Qy        41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       518 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 577

Qy        61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       578 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 637

Qy        81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
             |||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db       638 AACATCCCCAGCGTCCACGTGGGCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 697

Qy       101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       698 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 757
```

```
Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        758 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 817

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        818 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 877

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        878 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 937

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        938 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 997

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        998 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 1057

Qy        221 AspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGly 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1058 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 1117

Qy        241 LeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAspAsp 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1118 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 1177

Qy        261 Ala 261
              |||
Db       1178 GCC 1180
```

RESULT 6
ADF82458
ID ADF82458 standard; DNA; 2788 BP.
XX
AC ADF82458;
XX
DT 26-FEB-2004 (first entry)
XX
DE Leukaemia-related DNA sequence #3014.
XX
KW Cytostatic; Gene therapy; leukaemia; ss.
XX
OS Unidentified.
XX
PN WO2003039443-A2.
XX
PD 15-MAY-2003.
XX
PF 04-NOV-2002; 2002WO-EP012303.
XX
PR 05-NOV-2001; 2001EP-00126244.
PR 30-APR-2002; 2002EP-00009758.
XX
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA (UYLU-) UNIV LUDWIG MAXIMILIANS.
PA (HAFE/) HAFERLACH T.
PA (SCHO/) SCHOCH C.
PA (KERN/) KERN W.
XX
PI Haferlach T, Schoch C, Kern W, Kohlmann A, Schnittger S, Dugas M;
PI Eils R, Brors B, Mergenthaler S;
XX
DR WPI; 2003-505037/47.
XX
PT Determining the subtype of leukemia cells and whether a patient sample
PT contains leukemia cells or other cells, useful for treating leukemia,
PT comprises determining the expression profile of a group of markers in a
PT patient sample.
XX
PS Disclosure; SEQ ID NO 3014; 2938pp; English.
XX

CC The present invention relates to a method (M1) for determining the
CC subtype of leukaemia cells and whether a patient sample contains
CC leukaemia cells. The method comprises determining the expression profile
CC of a group of markers in a patient sample. The method is useful for
CC determining the presence of leukaemia cells, its types or subtypes, and
CC for the preparation of a medicament for treating leukaemia.
XX
SQ Sequence 2788 BP; 663 A; 764 C; 700 G; 595 T; 0 U; 66 Other;

Alignment Scores:

Pred. No.:	2.39e-109	Length:	2788
Score:	1345.00	Matches:	260
Percent Similarity:	99.62%	Conservative:	0
Best Local Similarity:	99.62%	Mismatches:	1
Query Match:	99.56%	Indels:	0
DB:	10	Gaps:	0

US-10-043-649-2 (1-261) x ADF82458 (1-2788)

```
Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 446

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 CAAGGCCANGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 506

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 566

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        567 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 626

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        627 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 686

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        687 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 746

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        747 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 806

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        807 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 866

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        867 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 926

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        927 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 986

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        987 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 1046

Qy        221 AspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGly 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1047 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 1106

Qy        241 LeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAspAsp 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1107 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 1166

Qy        261 Ala 261
              |||
```

```
Db       1167 GCC 1169

RESULT 7
AAC77202
ID   AAC77202 standard; cDNA; 837 BP.
XX
AC   AAC77202;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF2757 polynucleotide sequence SEQ ID NO:5513.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US008621.
XX
PR   31-MAR-1999;   99US-0127607P.
PR   02-APR-1999;   99US-0127636P.
PR   05-APR-1999;   99US-0127728P.
PR   30-MAR-2000; 2000US-00540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB42993.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease.
XX
PS   Claim 5; Page 4692-4693; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC   anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC   cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC   dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC   antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC   sequences can be used for determining the presence of or predisposition
CC   to, or preventing or treating pathological conditions associated with an
CC   ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC   proteins in gene therapy vectors. The proteins and nucleic acids may be
CC   used to treat cancers, proliferative disorders, neurodegenerative
CC   disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC   diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC   storage, systemic lupus erythematosus, severe combined immunodeficiency
CC   (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC   disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC   cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC   enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ   Sequence 837 BP; 176 A; 254 C; 245 G; 160 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.:                1.24e-103      Length:          837
Score:                    1273.00        Matches:         245
Percent Similarity:       99.60%         Conservative:    1
Best Local Similarity:    99.19%         Mismatches:      1
Query Match:              94.23%         Indels:          0
DB:                       3              Gaps:            0

US-10-043-649-2 (1-261) x AAC77202 (1-837)

Qy          15 SerLeuSerSerSerValGlnGlyGlnGlyProValThrMetGluAlaGluArgSerLys 34
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           4 AGCTTGAGTTCCTCTGTCCAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAG 63

Qy          35 AlaThrAlaValAlaLeuGlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArg 54
               |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db          64 GCCACAGCCGCGGCCCTGGGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGA 123

Qy          55 LeuGlyGluProLeuThrIleValSerGluAspGlyAspTrpTrpThrValLeuSerGlu 74
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         124 CTCGGGGAGCCATTGACCATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAA 183

Qy          75 ValSerGlyArgGluTyrAsnIleProSerValHisValAlaLysValSerHisGlyTrp 94
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         184 GTCTCAGGCAGAGAGTATAACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGG 243

Qy          95 LeuTyrGluGlyLeuSerArgGluLysAlaGluGluLeuLeuLeuProGlyAsnPro 114
               ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db         244 CTGTATGAGGGCCTGAGCAGGGAGAAAGCAGAGGACCTGCTGTTGTTACCTGGGAACCCT 303

Qy         115 GlyGlyAlaPheLeuIleArgGluSerGlnThrArgArgGlySerTyrSerLeuSerVal 134
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         304 GGAGGGGCCTTCCTCATCCGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTC 363

Qy         135 ArgLeuSerArgProAlaSerTrpAspArgIleArgHisTyrArgIleHisCysLeuAsp 154
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         364 CGCCTCAGCCGCCCTGCATCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGAC 423

Qy         155 AsnGlyTrpLeuTyrIleSerProArgLeuThrPheProSerLeuGlnAlaLeuValAsp 174
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         424 AATGGCTGGCTGTACATCTCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGAC 483

Qy         175 HisTyrSerGluLeuAlaAspAspIleCysCysLeuLeuLysGluProCysValLeuGln 194
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         484 CATTACTCTGAGCTGGCGGATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAG 543

Qy         195 ArgAlaGlyProLeuProGlyLysAspIleProLeuProValThrValGlnArgThrPro 214
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         544 AGGGCTGGCCCGCTCCCTGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCA 603

Qy         215 LeuAsnTrpLysGluLeuAspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGlu 234
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         604 CTCAACTGGAAAGAGCTGGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAG 663

Qy         235 SerLeuLeuSerGluGlyLeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGlu 254
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         664 TCTCTTCTCAGTGAGGGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAG 723

Qy         255 AlaValSerLeuAspAspAla 261
               |||||||||||||||||||||
Db         724 GCTGTCTCTTTGGATGATGCC 744

RESULT 8
ABQ99374
ID   ABQ99374 standard; cDNA; 1413 BP.
XX
AC   ABQ99374;
XX
DT   25-FEB-2003  (first entry)
XX
```

DE Human coding sequence SEQ ID 107.
XX
KW Human; expressed sequence tag; EST; chromosome 20;
KW haematopoietic disorder; central nervous system disease; viral infection;
KW peripheral nervous system disease; non-healing wound; infectious disease;
KW immune deficiency; immune disorder; bacterial infection; allergy; cancer;
KW fungal infection; autoimmune disorder; coagulation disorder; nootropic;
KW antiallergic; antiinflammatory; immunosuppressive; neuroprotective;
KW cytostatic; haemostatic; virucide; antibacterial; fungicide;
KW immunostimulant; cerebroprotective; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200259260-A2.
XX
PD 01-AUG-2002.
XX
PF 16-NOV-2001; 2001WO-US042950.
XX
PR 17-NOV-2000; 2000US-00714936.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;
PI Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR WPI; 2002-590824/63.
DR N-PSDB; ABP64788.
XX
PT New isolated polynucleotide, useful in research, diagnostic or
PT therapeutic methods, e.g. preventing or treating disorders involving
PT aberrant protein expression or biological activity.
XX
PS Claim 1; SEQ ID NO 107; 394pp; English.
XX
CC The present invention relates to novel human coding sequences (ABQ99268-
CC ABQ99608) and proteins (ABP64682-ABP65022). The sequences are useful in
CC therapeutic, diagnostic and research methods. The polynucleotides may be
CC used in the field of molecular biology as hybridisation probes, primers
CC for PCR, for chromosome and gene mapping, for the recombinant production
CC of protein, or in generation of anti-sense DNA or RNA. The
CC polynucleotides are useful in diagnostics as expressed sequence tags
CC (ESTs) for identifying expressed genes or for physical mapping of the
CC human genome. The proteins may be used as molecular weight markers, or as
CC nutritional sources or supplements. The proteins may be used to maintain
CC and expand cell population in a totipotential or pluripotential state
CC useful for re-engineering damaged or diseased tissues, transplantation,
CC manufacture of bio-pharmaceuticals or the development of bio-sensors. The
CC polynucleotides and proteins are useful for preventing, treating or
CC ameliorating disorders involving aberrant protein expression or
CC biological activity, e.g. haematopoietic disorders, central/peripheral
CC nervous system diseases, mechanical and traumatic disorders, non-healing
CC wounds, immune deficiencies and disorders, infectious diseases caused by
CC viral, bacterial or fungal infection, autoimmune disorders, allergic
CC reactions and conditions, coagulation disorders, or cancer. The
CC polynucleotide sequences of the invention were assembled from ESTs
CC isolated mainly by sequencing by hybridisation, and in some cases,
CC sequences obtained from one or more public databases. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1413 BP; 332 A; 407 C; 388 G; 286 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	8.64e-98	Length:	1413
Score:	1210.50	Matches:	241
Percent Similarity:	91.98%	Conservative:	0
Best Local Similarity:	91.98%	Mismatches:	3
Query Match:	89.60%	Indels:	18
DB:	6	Gaps:	1

US-10-043-649-2 (1-261) x ABQ99374 (1-1413)

```
Qy        1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
            |||||    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       54 ATGGGAGGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 113

Qy       21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      114 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 173

Qy       41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      174 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 233

Qy       61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      234 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 293

Qy       81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      294 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 353

Qy      101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      354 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 413

Qy      121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      414 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 473

Qy      141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      474 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 533

Qy      161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
            ||||||||||||||||||||||||||||||||||   |||||||||
Db      534 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGGGGACCATTAC------------ 581

Qy      181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGln-ArgAlaGlyProLeuPr 200
                                                       ||||||||||||||||
Db      582 ----------------------------------------TCTGAGGGCTGGCCCGCTCCC 602

Qy      200 oGlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLe 220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      603 TGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCT 662

Qy      220 uAspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGl 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      663 GGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGG 722

Qy      240 yLeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAspAs 260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      723 TCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGA 782

Qy      260 pAla 261
            ||||
Db      783 TGCC 786
```

RESULT 9
AAL44090
ID AAL44090 standard; cDNA; 737 BP.
XX
AC AAL44090;
XX
DT 03-OCT-2002 (first entry)
XX
DE Mouse MARS short isoform protein coding sequence.
XX
KW Mouse; gene; ss; gene therapy; modulator of antigen receptor signalling;
KW MARS; tumour suppressor gene; Scr-like adaptor protein; SLAP;
KW myeloid malignancy; acute myelogenous leukaemia; autoimmune disorder;
KW immunosuppression; myeloproliferative disorder; breast cancer.
XX
OS Mus sp.

```
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .633
FT                   /*tag=  a
FT                   /product= "Mouse MARS short isoform protein"
XX
PN   WO200242452-A2.
XX
PD   30-MAY-2002.
XX
PF   26-NOV-2001; 2001WO-CA001662.
XX
PR   27-NOV-2000; 2000CA-02324663.
XX
PA   (HOSP-) HOSPITAL FOR SICK CHILDREN.
XX
PI   Mcglade JC,  Loreto MP;
XX
DR   WPI; 2002-566564/60.
DR   P-PSDB; AAO15458.
XX
PT   New isolated modulator of antigen receptor signaling protein or its
PT   fragment, useful for treating malignant disorders such as myeloid
PT   malignancies, autoimmune disorders and myeloproliferative disorders.
XX
PS   Claim 9; Page 77; 110pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of modulator
CC   of antigen receptor signalling (MARS) proteins. The MARS protein is a
CC   putative tumour suppressor gene and exhibits structural and sequence
CC   similarity to the Scr-like adaptor protein (SLAP). The MARS DNA and
CC   protein sequences of the invention are useful for the treatment of
CC   myeloid malignancies (e.g. acute myelogenous leukaemia) autoimmune
CC   disorders, immunosuppression, myeloproliferative disorders and
CC   malignancies related to the de-regulation of tyrosine kinases (e.g.
CC   breast cancer). The present cDNA sequence encodes a mouse MARS protein
XX
SQ   Sequence 737 BP; 152 A; 219 C; 218 G; 148 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                2.87e-97      Length:          737
Score:                    1200.50       Matches:         241
Percent Similarity:       92.40%        Conservative:    2
Best Local Similarity:    91.63%        Mismatches:      1
Query Match:              88.86%        Indels:          19
DB:                       6             Gaps:            1

US-10-043-649-2 (1-261) x AAL44090 (1-737)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerServal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTAC------------ 528

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGln-ArgAlaGlyProLeuPr 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        529 ----------------------------------------TCTGAGGGCTGGCCCGCTCCC 549

Qy        200 oGlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLe 220
              |||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        550 TGGCAAGGATATACCCCTACCTGTGACTGTGCGGAGGACACCACTCAACTGGAAAGAGCT 609

Qy        220 uAspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGl 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        610 GGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGG 669

Qy        240 yLeuArgGluSerLeuSerPheTyrIleSerLeuAsnAsp-GluAlaValSerLeuAspA 260
              |||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||||
Db        670 TCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGAGCGAGGCTGTCTCTTTGGATG 729

Qy        260 spAla 261
              |||||
Db        730 ATGCC 734

RESULT 10
AAS74750
ID   AAS74750 standard; cDNA; 2049 BP.
XX
AC   AAS74750;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #10554.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG10563.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 10554; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
```

```
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2049 BP; 479 A; 573 C; 551 G; 443 T; 0 U; 3 Other;

Alignment Scores:
Pred. No.:              1.18e-90      Length:        2049
Score:                  1132.50       Matches:       239
Percent Similarity:     89.89%        Conservative:  1
Best Local Similarity:  89.51%        Mismatches:    4
Query Match:            83.83%        Indels:        23
DB:                     5             Gaps:          1

US-10-043-649-2 (1-261) x AAS74750 (1-2049)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        965 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 1024

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1025 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 1084

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1085 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 1144

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1145 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 1204

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1205 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 1264

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1265 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 1324

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              |||||||||||||||||||||||||||||||||||||||||||||:::|||||||||||
Db       1325 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAACCGCCCTGCA 1384

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              |||||||||||||||||||||||||||   |||||||||||||||||||||||||||||
Db       1385 TTCTGGGACCGGATCAGACACTACAGGATTCACTGGCTTGACAATGGCTGGCTGTACATC 1444

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1445 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTAC------------ 1492

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGln-ArgAlaGlyProLeuPr 200
                                                       ||||||||||||||||||
Db       1493 -----------------------------------TCTGAGGGCTGGCCCGCTCCC 1513

Qy        200 oGlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLe 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       1514 TGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCT 1573

Qy        220 uAspSerSerLeuLeuPheSerGluAlaAlaThr-GlyGluGluSerLeuLeu-SerGlu 239
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||     |||
Db       1574 GGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGGAGGAGTCTCTTCTCAGAGGAG 1633

Qy        240 GlyLeuArgGluSer-LeuSerPheTyrIleSer-LeuAsnAspGluAlaValSer-Leu 258
              |||||||||||||||| |||||||||||||||||| |||||||||||||||||||| ||| 
Db       1634 GGGCTCCGGGAGTCCCCTCAGCTTCTACATCAGCCCTGAATGACGAGGCTGTCTCTTTTG 1693

Qy        259 AspAspAla 261
              |||||||||
Db       1694 GATGATGCC 1702

RESULT 11
AAL44087
ID   AAL44087 standard; cDNA; 1348 BP.
XX
AC   AAL44087;
XX
DT   03-OCT-2002  (first entry)
XX
DE   Mouse modulator of antigen receptor signalling protein coding sequence.
XX
KW   Mouse; gene; ss; gene therapy; modulator of antigen receptor signalling;
KW   MARS; tumour suppressor gene; Scr-like adaptor protein; SLAP;
KW   myeloid malignancy; acute myelogenous leukaemia; autoimmune disorder;
KW   immunosuppression; myeloproliferative disorder; breast cancer.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             282. .1061
FT                   /*tag=  a
FT                   /product= "Mouse MARS protein"
XX
PN   WO200242452-A2.
XX
PD   30-MAY-2002.
XX
PF   26-NOV-2001; 2001WO-CA001662.
XX
PR   27-NOV-2000; 2000CA-02324663.
XX
PA   (HOSP-) HOSPITAL FOR SICK CHILDREN.
XX
PI   Mcglade JC,  Loreto MP;
XX
DR   WPI; 2002-566564/60.
DR   P-PSDB; AAO15456.
XX
PT   New isolated modulator of antigen receptor signaling protein or its
PT   fragment, useful for treating malignant disorders such as myeloid
PT   malignancies, autoimmune disorders and myeloproliferative disorders.
XX
PS   Claim 10; Fig 1A; 110pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of modulator
CC   of antigen receptor signalling (MARS) proteins. The MARS protein is a
CC   putative tumour suppressor gene and exhibits structural and sequence
CC   similarity to the Scr-like adaptor protein (SLAP). The MARS DNA and
CC   protein sequences of the invention are useful for the treatment of
CC   myeloid malignancies (e.g. acute myelogenous leukaemia) autoimmune
CC   disorders, immunosuppression, myeloproliferative disorders and
CC   malignancies related to the de-regulation of tyrosine kinases (e.g.
CC   breast cancer). The present cDNA sequence encodes a mouse MARS protein
XX
SQ   Sequence 1348 BP; 324 A; 385 C; 362 G; 277 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              5.73e-82      Length:        1348
Score:                  1032.00       Matches:       209
Percent Similarity:     85.88%        Conservative:  16
```

```
Best Local Similarity:  79.77%      Mismatches:    33
Query Match:            76.39%      Indels:         4
DB:                     6           Gaps:           3

US-10-043-649-2 (1-261) x AAL44087 (1-1348)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              ||||||||||||  ||||||  |||:::      |||||||||  |||||||||
Db        282 ATGGGAAGTTTGTCCAGCAGAGGGAAAACC---TCCAGCCCCAGCCCCAGCTCCTCTGGT 338

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
                 |||   ||||||:::|||:::   ||||||  |||   ||||||||||||||||
Db        339 CCAGACCAGGAACCCGTGTCCATGCAACCAGAAAGACACAAGGTCACAGCTGTGGCCCTG 398

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||       |||   ||||||||||||||||||||||||||||
Db        399 GGCAGTTTCCCAGCAGGTGAACAGGCCAGACTATCTCTGAGACTCGGGGAGCCGCTGACC 458

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              |||:::|||||||||||||||||||||||||   ||||||||||||||||||||||||
Db        459 ATCATCTCTGAGGATGGAGATTGGTGGACAGTCCAGTCGGAAGTCTCAGGCAGAGAGTAC 518

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ::::::|||||||||:::|||||||||||||:::|||||||||||||||||||||||||
Db        519 CACATGCCCAGTGTGTATGTGGCTAAAGTCGCCCACGGGTGGCTGTACGAGGGCCTGAGC 578

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        579 CGGGAGAAAGCCGAGGAACTACTCCTGTTACCTGGGAACCCCGGAGGGGCCTTCCTCATC 638

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              |||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db        639 CGGGAGAGCCAGACCAGGAGAGGCTGCTATTCCCTGTCCGTCCGACTCAGCCGCCCTGCA 698

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              |||||||||||||||||||||||||||||      ||||||||||||||||||||||||
Db        699 TCTTGGGACCGGATCAGACACTACAGGATACAGCGTCTTGACAATGGCTGGCTGTACATC 758

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              |||||||||||||||||||||||||||   |||||||||:::|||||||||||||||||
Db        759 TCACCTCGCCTCACCTTCCCCTCACTCCACGCCTTGGTGGAGCATTACTCTGAGCTAGCA 818

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              |||   |||||||||   |||:::|||||||||||||||||:::   ||||||||||||
Db        819 GATGGCATCTGCTGTCCCCTCAGGGAGCCGTGTGTCCTGCAGAAGCTTGGGCCACTACCT 878

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              ||||||||   |||   ||||||||||||      :::   ||||||||||||:::|||
Db        879 GGCAAAGATACACCTCCACCTGTGACTGTGCCAACATCATCACTAAATTGGAAAAAGCTG 938

Qy        221 AspSerSerLeuLeuPheSerGluAla---AlaThrGlyGluGluSerLeuLeuSerGlu 239
              |||   ||||||||||||   ||||||   |||:::||||||   ||||||||||||||
Db        939 GACCGCAGCCTCCTGTTTCTGGAAGCACCTGCGAGTGGGGAGGCATCTCTGCTCAGTGAG 998

Qy        240 GlyLeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAsp 259
              |||||||||||||||||||   ||||||||||||   ::::::          ||||||
Db        999 GGGCTCCGAGAGTCCCTCAGTTCCTACATCAGCCTGGCTGAGGAC------CCCTTGGAT 1052

Qy        260 AspAla 261
              ||||||
Db       1053 GATGCT 1058

RESULT 12
ABQ98670
ID   ABQ98670 standard; DNA; 763 BP.
XX
AC   ABQ98670;
XX
DT   04-NOV-2002  (first entry)
XX
DE   Human ORF477 coding sequence.
XX
KW   Cytostatic; Cardiant; Anti-allergic; Immunosuppressive; Vulnerary;
KW   Antiinflammatory; gene therapy; human; ORFX; atherogenic; platelet;
KW   human umbilical vein endothelial cell; HUVEC; atherosclerotic plaque;
KW   cancer; cardiovascular disease; allergy; autoimmune disease;
KW   wound healing; blood coagulation disorder; inflammatory disorder; ds.
XX
OS   Homo sapiens.
XX
PN   US2002082206-A1.
XX
PD   27-JUN-2002.
XX
PF   30-MAY-2001; 2001US-00867550.
XX
PR   30-MAY-2000; 2000US-0208427P.
XX
PA   (LEAC/) LEACH M D.
PA   (MEHR/) MEHRABAN F.
PA   (CONL/) CONLEY P B.
PA   (TOPP/) TOPPER J N.
PA   (LAWD/) LAW D.
XX
PI   Leach MD,  Mehraban F,  Conley PB,  Topper JN,  Law D;
XX
DR   WPI; 2002-626554/67.
DR   P-PSDB; ABP64107.
XX
PT   New polypeptide designated ORFX are present in human atherogenic cells
PT   and are useful to prevent and treat ORFX-associated disorders including
PT   cancer, allergy, wound healing or autoimmune, cardiovascular or
PT   inflammatory disease.
XX
PS   Claim 2; SEQ ID NO 953; 78pp; English.
XX
CC   The present invention relates to novel human ORFX polypeptides and their
CC   coding sequences (ABP63631-ABP64681 and ABQ98194-ABQ99267). The sequences
CC   were discovered in human atherogenic cells, in particular in platelets
CC   and human umbilical vein endothelial cells (HUVEC) and are expressed in
CC   many other tissues as well. Atherogenic cells are cells which have the
CC   potential to develop atherosclerotic plaques. The ORFX polypeptides and
CC   nucleic acids are useful for treating or preventing a pathological
CC   condition associated with an ORFX-associated disorder, e.g. cancer,
CC   cardiovascular disease, allergy, autoimmune disease, wound healing, blood
CC   coagulation disorders or inflammatory disorders. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from the USPTO web site at
CC   seqdata.uspto.gov/sequence.html?DocID=20020082206
XX
SQ   Sequence 763 BP; 176 A; 222 C; 218 G; 147 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.36e-64      Length:         763
Score:                  830.00        Matches:        159
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            61.44%        Indels:         0
DB:                     6             Gaps:           0

US-10-043-649-2 (1-261) x ABQ98670 (1-763)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        286 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 345

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        346 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 405

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        406 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 465

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        466 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 525

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        526 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 585

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        586 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 645

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        646 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 705

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyr 159
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        706 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTAC 762
```

RESULT 13
AAS74748
ID AAS74748 standard; cDNA; 603 BP.
XX
AC AAS74748;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #10552.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG10561.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 10552; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 603 BP; 124 A; 189 C; 164 G; 126 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.07e-55	Length:	603
Score:	731.00	Matches:	151
Percent Similarity:	82.45%	Conservative:	4
Best Local Similarity:	80.32%	Mismatches:	21
Query Match:	54.11%	Indels:	12
DB:	5	Gaps:	4

US-10-043-649-2 (1-261) x AAS74748 (1-603)

```
Qy         78 ArgGluTyrAsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGlu 97
              |||       |||:::|||         |||   |||:::          |||||||||
Db         61 CGCAGGGGGAACCTGCCCAAGGAGTCGGTG---AAGATCCTCCGGGACTGGCTGTAC--- 114

Qy         98 GlyLeuSerArgGluLysAla-----------GluGluLeuLeuLeuProGlyAsn 113
                 |||   |||       |||          :::|||   |||   |||   |||
Db        115 ---TTGCACCGCTACAACGCCTACCCCTCAGAGCAGGAGAAGCTGAGCCTTTCTGGACAG 171

Qy        114 ProGlyGlyAlaPheLeuIleArgGluSerGlnThrArgArgGlySerTyrSerLeuSer 133
                    :::   |||                |||   ||||||||||||||||||
Db        172 ACCAACCTGTCAGTGCTG--------------CAAAGACATGGCTCTTACTCTCTGTCA 216

Qy        134 ValArgLeuSerArgProAlaSerTrpAspArgIleArgHisTyrArgIleHisCysLeu 153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        217 GTCCGCCTCAGCCGCCCTGCATCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTT 276

Qy        154 AspAsnGlyTrpLeuTyrIleSerProArgLeuThrPheProSerLeuGlnAlaLeuVal 173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        277 GACAATGGCTGGCTGTACATCTCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTG 336

Qy        174 AspHisTyrSerGluLeuAlaAspAspIleCysCysLeuLeuLysGluProCysValLeu 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        337 GACCATTACTCTGAGCTGGCGGATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTG 396

Qy        194 GlnArgAlaGlyProLeuProGlyLysAspIleProLeuProValThrValGlnArgThr 213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        397 CAGAGGGCTGGCCCGCTCCCTGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACA 456

Qy        214 ProLeuAsnTrpLysGluLeuAspSerSerLeuLeuPheSerGluAlaAlaThrGlyGlu 233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        457 CCACTCAACTGGAAAGAGCTGGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAG 516

Qy        234 GluSerLeuLeuSerGluGlyLeuArgGluSerLeuSerPheTyrIleSerLeuAsnAsp 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        517 GAGTCTCTTCTCAGTGAGGGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGAC 576

Qy        254 GluAlaValSerLeuAspAspAla 261
              |||||||||||||||||||||||
Db        577 GAGGCTGTCTCTTTGGATGATGCC 600
```

RESULT 14
ADL63090
ID ADL63090 standard; DNA; 864 BP.
XX
AC ADL63090;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human ovarian cancer DNA marker #21302.
XX
KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
XX
OS Homo sapiens.
XX
PN WO200170979-A2.
XX

PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009126.
XX
PR 21-MAR-2000; 2000US-0191031P.
PR 25-MAY-2000; 2000US-0207124P.
PR 15-JUN-2000; 2000US-0211940P.
PR 07-JUL-2000; 2000US-0216820P.
PR 25-JUL-2000; 2000US-0220661P.
PR 21-DEC-2000; 2000US-0257672P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lee J, Lillie J;
XX
DR WPI; 2001-611502/70.
XX
PT Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
PT cancer cells as compared to their normal non-cancerous ovarian cells are
PT used to characterize stage, grade, histological type of ovarian cancer.
XX
PS Disclosure; SEQ ID NO 21302; 106pp; English.
XX
CC The invention relates to nucleic acid markers which are overexpressed in
CC ovarian cancer cells as compared to their expression in normal (i.e. non-
CC cancerous) ovarian cells. The invention also relates to polypeptides
CC encoded by the markers, antibodies that selectively bind to the
CC polypeptides, a method of inhibiting ovarian cancer in a patient at risk
CC of developing ovarian cancer involving inhibiting expression of a gene
CC corresponding to a marker of the invention and a method of treating a
CC patient afflicted with ovarian cancer comprising providing to cells of
CC the patient an antisense oligonucleotide complementary to a marker of the
CC invention. The markers are useful for assessing if a patient is afflicted
CC with ovarian cancer, which involves comparing the level of expression of
CC a marker in a patient sample and a normal level of expression of the
CC marker in a control non-ovarian cancer sample. A difference between the
CC expression levels indicates ovarian cancer. The level of expression of a
CC marker corresponds to a secreted protein or to a transcribed
CC polynucleotide or its portion. The level of expression of the marker is
CC assessed by detecting the presence in the sample, a protein or protein
CC fragment corresponding to the marker. The presence of protein or protein
CC fragment is detected using an antibody that specifically binds with the
CC protein or protein fragment. Alternatively, the level of expression of
CC the marker is assessed by detecting the presence of a transcribed
CC polynucleotide which anneals with the marker or anneals with a portion of
CC the polynucleotide comprising the marker, under stringent conditions. The
CC marker is also used for monitoring the progression of ovarian cancer in a
CC patient which involves detecting expression of the marker in a patient
CC sample at a first point in time, repeating the method at a subsequent
CC time and comparing the level of expression. The method is carried out
CC using an ovarian tissue sample. A composition comprising a marker,
CC polypeptide or antibody of the invention is used to treat ovarian cancer.
CC This sequence represents a human ovarian cancer DNA marker of the
CC invention.
XX
SQ Sequence 864 BP; 191 A; 259 C; 242 G; 165 T; 0 U; 7 Other;

Alignment Scores:

Pred. No.:	4.9e-48	Length:	864
Score:	647.00	Matches:	127
Percent Similarity:	99.22%	Conservative:	1
Best Local Similarity:	98.45%	Mismatches:	1
Query Match:	47.89%	Indels:	0
DB:	5	Gaps:	0

US-10-043-649-2 (1-261) x ADL63090 (1-864)

```
Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerServal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        450 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 509

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        510 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 569

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        570 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 629

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        630 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 689

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        690 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 749

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        750 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 809

Qy        121 ArgGluSerGlnThrArgArgGlySer 129
              |||||||||||||||||:::   |||
Db        810 CGGGAGAGCCAGACCAGGAAGAGGTCC 836
```

RESULT 15
ABQ99151
ID ABQ99151 standard; DNA; 875 BP.
XX
AC ABQ99151;
XX
DT 04-NOV-2002 (first entry)
XX
DE Human ORF958 coding sequence.
XX
KW Cytostatic; Cardiant; Anti-allergic; Immunosuppressive; Vulnerary;
KW Antiinflammatory; gene therapy; human; ORFX; atherogenic; platelet;
KW human umbilical vein endothelial cell; HUVEC; atherosclerotic plaque;
KW cancer; cardiovascular disease; allergy; autoimmune disease;
KW wound healing; blood coagulation disorder; inflammatory disorder; ds.
XX
OS Homo sapiens.
XX
PN US2002082206-A1.
XX
PD 27-JUN-2002.
XX
PF 30-MAY-2001; 2001US-00867550.
XX
PR 30-MAY-2000; 2000US-0208427P.
XX
PA (LEAC/) LEACH M D.
PA (MEHR/) MEHRABAN F.
PA (CONL/) CONLEY P B.
PA (TOPP/) TOPPER J N.
PA (LAWD/) LAW D.
XX
PI Leach MD, Mehraban F, Conley PB, Topper JN, Law D;
XX
DR WPI; 2002-626554/67.
DR P-PSDB; ABP64588.
XX
PT New polypeptide designated ORFX are present in human atherogenic cells
PT and are useful to prevent and treat ORFX-associated disorders including
PT cancer, allergy, wound healing or autoimmune, cardiovascular or
PT inflammatory disease.
XX
PS Claim 2; SEQ ID NO 1915; 78pp; English.
XX
CC The present invention relates to novel human ORFX polypeptides and their
CC coding sequences (ABP63631-ABP64681 and ABQ98194-ABQ99267). The sequences
CC were discovered in human atherogenic cells, in particular in platelets
CC and human umbilical vein endothelial cells (HUVEC) and are expressed in
CC many other tissues as well. Atherogenic cells are cells which have the
CC potential to develop atherosclerotic plaques. The ORFX polypeptides and
CC nucleic acids are useful for treating or preventing a pathological

CC condition associated with an ORFX-associated disorder, e.g. cancer,
CC cardiovascular disease, allergy, autoimmune disease, wound healing, blood
CC coagulation disorders or inflammatory disorders. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from the USPTO web site at
CC seqdata.uspto.gov/sequence.html?DocID=20020082206
XX
SQ Sequence 875 BP; 205 A; 259 C; 225 G; 185 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.:	1.3e-42	Length:	875
Score:	586.00	Matches:	112
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	43.38%	Indels:	0
DB:	6	Gaps:	0

US-10-043-649-2 (1-261) x ABQ99151 (1-875)

```
Qy        150 IleHisCysLeuAspAsnGlyTrpLeuTyrIleSerProArgLeuThrPheProSerLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 ATCCACTGCCTTGACAATGGCTGGCTGTACATCTCACCGCGCCTCACCTTCCCCTCACTC 63

Qy        170 GlnAlaLeuValAspHisTyrSerGluLeuAlaAspAspIleCysCysLeuLeuLysGlu 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 CAGGCCCTGGTGGACCATTACTCTGAGCTGGCGGATGACATCTGCTGCCTACTCAAGGAG 123

Qy        190 ProCysValLeuGlnArgAlaGlyProLeuProGlyLysAspIleProLeuProValThr 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 CCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCTGGCAAGGATATACCCCTACCTGTGACT 183

Qy        210 ValGlnArgThrProLeuAsnTrpLysGluLeuAspSerSerLeuLeuPheSerGluAla 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 GTGCAGAGGACACCACTCAACTGGAAAGAGCTGGACAGCTCCCTCCTGTTTTCTGAAGCT 243

Qy        230 AlaThrGlyGluGluSerLeuLeuSerGluGlyLeuArgGluSerLeuSerPheTyrIle 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 GCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGTCTCCGGGAGTCCCTCAGCTTCTACATC 303

Qy        250 SerLeuAsnAspGluAlaValSerLeuAspAspAla 261
              ||||||||||||||||||||||||||||||||||||
Db        304 AGCCTGAATGACGAGGCTGTCTCTTTGGATGATGCC 339
```

Search completed: November 17, 2004, 00:59:18
Job time : 539 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 16, 2004, 11:56:01 ; Search time 509 Seconds
(without alignments)
8106.179 Million cell updates/sec

Title: US-10-043-649-1
Perfect score: 786
Sequence: 1 atgggaagtctgcccagcag.....tctctttggatgatgcctag 786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	786	100.0	786	6	AAL44089	Aal44089 Human mod
2	786	100.0	786	6	ABQ74343	Abq74343 Human Src
3	786	100.0	2567	6	AAD43980	Aad43980 Human Src
c 4	786	100.0	3646	10	ADF90741	Adf90741 Human hep
5	785	99.9	2788	10	ADF82458	Adf82458 Leukaemia
6	784.4	99.8	1183	6	ABK61465	Abk61465 Human cDN
7	741.8	94.4	837	3	AAC77202	Aac77202 Human ORF
8	672.8	85.6	1413	6	ABQ99374	Abq99374 Human cod
9	663.4	84.4	737	6	AAL44090	Aal44090 Mouse MAR
10	609.8	77.6	2049	5	AAS74750	Aas74750 DNA encod
11	537.8	68.4	1348	6	AAL44087	Aal44087 Mouse mod
12	478	60.8	763	6	ABQ98670	Abq98670 Human ORF
13	405	51.5	603	5	AAS74748	Aas74748 DNA encod
14	381	48.5	864	5	ADL63090	Adl63090 Human ova
15	341	43.4	875	6	ABQ99151	Abq99151 Human ORF
16	215	27.4	320	5	ADL43424	Adl43424 Human ova
17	192.8	24.5	211	5	AAS70181	Aas70181 DNA encod
18	192	24.4	444	6	ABQ98669	Abq98669 Human ORF
19	160.2	20.4	2109	4	AAS02049	Aas02049 DNA encod
20	159	20.2	2665	6	ABL65189	Abl65189 Lung canc
21	159	20.2	2665	6	ABK83738	Abk83738 Human cDN
22	159	20.2	2665	8	ACC81091	Acc81091 Human Src
23	159	20.2	2665	12	ADL83115	Adl83115 Human PRO
24	159	20.2	2665	12	ADP12833	Adp12833 Reference
25	159	20.2	3452	12	ADQ22520	Adq22520 Human sof
26	159	20.2	3756	8	ABX62975	Abx62975 Human act
27	141.8	18.0	432	4	AAI12879	Aai12879 Probe #28
28	141.8	18.0	432	4	ABA54580	Aba54580 Human foe
29	141.8	18.0	432	4	AAI34236	Aai34236 Probe #29
30	141.8	18.0	432	4	ABA44128	Aba44128 Human bre
31	141.8	18.0	432	4	ABA24363	Aba24363 Probe #28
32	141.8	18.0	432	4	AAK28314	Aak28314 Human bon
33	141.8	18.0	432	4	AAK02872	Aak02872 Human bra
34	141.8	18.0	432	4	ABS27912	Abs27912 Human liv
35	141.8	18.0	432	5	AAI02797	Aai02797 Probe #27
36	141.8	18.0	432	6	ABS02823	Abs02823 Human gen
37	141.8	18.0	448	4	AAI14520	Aai14520 Probe #44
38	141.8	18.0	448	6	ABS04499	Abs04499 Human gen
39	133.2	16.9	152	5	ADI71889	Adi71889 Human ova
40	133.2	16.9	152	5	ADL37040	Adl37040 Human ova
41	124.6	15.9	1924	12	ADI04091	Adi04091 Human HCK
42	124.6	15.9	1926	6	ABK83940	Abk83940 Human cDN
43	124.6	15.9	1926	12	ADJ71658	Adj71658 Human NOV
44	124.6	15.9	2015	6	ABL66673	Abl66673 Lung canc
45	124.6	15.9	2015	6	ABK83939	Abk83939 Human cDN

ALIGNMENTS

RESULT 1
AAL44089
ID AAL44089 standard; cDNA; 786 BP.
XX
AC AAL44089;
XX
DT 03-OCT-2002 (first entry)
XX
DE Human modulator of antigen receptor signalling protein coding sequence.
XX
KW Human; gene; ss; gene therapy; modulator of antigen receptor signalling;
KW MARS; tumour suppressor gene; Scr-like adaptor protein; SLAP;
KW myeloid malignancy; acute myelogenous leukaemia; autoimmune disorder;
KW immunosuppression; myeloproliferative disorder; breast cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1. .786
FT /*tag= a
FT /product= "Human MARS protein"
XX
PN WO200242452-A2.
XX
PD 30-MAY-2002.
XX
PF 26-NOV-2001; 2001WO-CA001662.
XX
PR 27-NOV-2000; 2000CA-02324663.
XX
PA (HOSP-) HOSPITAL FOR SICK CHILDREN.
XX
PI Mcglade JC, Loreto MP;
XX
DR WPI; 2002-566564/60.
DR P-PSDB; AAO15457.
XX
PT New isolated modulator of antigen receptor signaling protein or its
PT fragment, useful for treating malignant disorders such as myeloid
PT malignancies, autoimmune disorders and myeloproliferative disorders.
XX
PS Claim 12; Page 75; 110pp; English.
XX
CC The invention comprises the amino acid and coding sequences of modulator

```
CC   of antigen receptor signalling (MARS) proteins. The MARS protein is a
CC   putative tumour suppressor gene and exhibits structural and sequence
CC   similarity to the Scr-like adaptor protein (SLAP). The MARS DNA and
CC   protein sequences of the invention are useful for the treatment of
CC   myeloid malignancies (e.g. acute myelogenous leukaemia) autoimmune
CC   disorders, immunosuppression, myeloproliferative disorders and
CC   malignancies related to the de-regulation of tyrosine kinases (e.g.
CC   breast cancer). The present cDNA sequence encodes a human MARS protein
XX
SQ   Sequence 786 BP; 162 A; 234 C; 231 G; 159 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 786;  DB 6;  Length 786;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-197;
  Matches  786;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATGACATCTGCTGCCTACTCAAGGAGCCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780

Qy        781 GCCTAG 786
              ||||||
Db        781 GCCTAG 786
```

RESULT 2

```
ABQ74343
ID   ABQ74343 standard; cDNA; 786 BP.
XX
AC   ABQ74343;
XX
DT   15-OCT-2002  (first entry)
XX
DE   Human Src-like inhibitory molecule (SLIM) encoding cDNA.
XX
KW   Human; Src-like inhibitory molecule; SLIM; Src-like adapter protein;
KW   SLAP; inhibitor; antiinflammatory; immunosuppressive; anti-HIV;
KW   modulator; lymphocyte; Cbl; gene therapy; immunodeficiency disorder;
KW   acquired immune deficiency syndrome; AIDS; acute inflammatory disorder;
KW   chronic inflammatory disorder; autoimmune disorder; transplant rejection;
KW   gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .786
FT                   /*tag=  a
FT                   /product= "SLIM"
FT                   /note= "Src-like inhibitory molecule"
XX
PN   WO200255707-A2.
XX
PD   18-JUL-2002.
XX
PF   10-JAN-2002; 2002WO-US000718.
XX
PR   10-JAN-2001; 2001US-0260953P.
XX
PA   (RIGE-) RIGEL PHARM INC.
XX
PI   Holland SJ,  Mendenhall MK,  Pardo J,  Spencer C,  Fu AC,  Luo Y;
PI   Payan DG,  Mancebo HS,  Wu J,  Zhou X,  Shen M,  Liao XC,  Sheng N;
XX
DR   WPI; 2002-575432/61.
DR   P-PSDB; ABP52187.
XX
PT   New src-like inhibitory molecule protein, useful for treating
PT   immunodeficiency disorders and inflammatory disorders, comprises N-
PT   terminal myristylation sequence, SH2 domain and/or SH3 domain.
XX
PS   Claim 3; Fig 2A; 91pp; English.
XX
CC   The present sequence encodes the human Src-like inhibitory molecule
CC   (SLIM) protein (I). The present invention describes a SLIM protein
CC   comprising an N-terminal myristylation sequence, an N-terminal SH2
CC   domain, and an N-terminal SH3 domain which can bind to Cbl, or comprising
CC   an N-terminal myristylation sequence and an N-terminal SH2 domain which
CC   is unable to bind to Cbl. (I) has antiinflammatory, immunosuppressive and
CC   anti-HIV activities, and can be used as a modulator of lymphocyte
CC   activation, and of ubiquitination of a Cbl target protein, and in gene
CC   therapy. (I) is useful for screening a bioactive agent capable of binding
CC   to SLIM. (I) is also useful for screening a bioactive agent capable of
CC   modulating SLIM binding. (I) or its fragments is useful in the study or
CC   in the treatment of conditions which involves this function or
CC   dysregulation of SLIM protein activity, i.e. to diagnose, treat or
CC   prevent SLIM associated disorders. (I) or the polynucleotide encoding it
CC   (II) is useful for modulating leukocyte and/or platelet activation, for
CC   modulating antigen receptor-induced signalling and activation in
CC   leukocyte and/or platelets and for modulating antigen receptor-induced
CC   signalling and activation in lymphocytes and/or mast cells. (I) or (II)
CC   is also useful for modulating the basal activity of lymphocytes. (I) or
CC   (II) is useful in the treatment of immunodeficiency disorders, such as
CC   acquired immunodeficiency syndrome (AIDS), for the prevention and
CC   treatment of acute inflammatory disorders, chronic inflammatory
CC   disorders, autoimmune disorder and transplant rejection
XX
SQ   Sequence 786 BP; 162 A; 234 C; 231 G; 159 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 786;  DB 6;  Length 786;
```

```
  Best Local Similarity   100.0%;  Pred. No. 1.1e-197;
  Matches  786;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480

Qy        481 TCACCGCGCCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCACCGCGCCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780

Qy        781 GCCTAG 786
              ||||||
Db        781 GCCTAG 786
```

RESULT 3
AAD43980
ID AAD43980 standard; cDNA; 2567 BP.
XX
AC AAD43980;
XX
DT 13-DEC-2002 (first entry)
XX
DE Human Src-Like Adapter Protein-2 (hSLAP-2) cDNA.
XX
KW Human; SH2/SH3-domain-containing adapter; Src-Like Adapter Protein-2;
KW SLAP-2; immune disorder; signal transduction; autoimmune disease; cancer;
KW neoplasm; inflammation disorder; rheumatoid arthritis; osteoarthritis;
KW psoriasis; rhinitis; allergy; inflammatory bowel disease; gene therapy;
KW Crohn's disease; systemic lupus erythematosus; tissue/organ rejection;
KW multiple sclerosis; asthma; acute respiratory distress syndrome;
KW pulmonary disorder; dermatological; neuroprotective; gene; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   415. .1200
FT   /*tag= a
FT   /product= "Human SLAP-2"
XX
PN WO200242457-A1.
XX
PD 30-MAY-2002.
XX
PF 20-NOV-2001; 2001WO-US043367.
XX
PR 22-NOV-2000; 2000US-0252545P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Chang H, Yang W, Wu Y, Whitney GS, Perez-Villar JJ, Kanner SB;
XX
DR WPI; 2002-463632/49.
DR P-PSDB; AAE26357.
XX
PT Novel substantially purified human SH2/SH3-domain-containing adapter
PT polypeptide, termed Src-Like Adapter Protein-2, useful for therapeutic
PT intervention in immunological and inflammatory disorders and cancer.
XX
PS Claim 2; Fig 1; 85pp; English.
XX
CC The invention relates to a substantially purified human SH2/SH3-domain-
CC containing adapter polypeptide, termed Src-Like Adapter Protein-2 (SLAP-
CC 2). The invention is useful for treating an immune disorder involving
CC hyperactivity of B- or T- lymphocytes in a mammal. The invention is
CC useful for screening for antagonists or inhibitors of the interaction of
CC hSLAP-2 with cellular signalling compounds, for diagnosing, treating or
CC preventing diseases or disorders associated with aberrant or uncontrolled
CC cellular signal transduction, for determining those cellular signalling
CC molecules which associate with hSLAP-2 and which provide critical signals
CC for cell activation, and as effectors in methods to affect T- cell
CC activation. The invention is useful in screening assays to identity and
CC detect candidate bioactive agents that modulate hSLAP-2 bioactivity, for
CC potential use to treat autoimmune diseases which may be caused by
CC hyperactivated B cells, as well as to treat diseases which may be caused
CC by hyperactivated T cells, in addition to other immune system related
CC conditions, diseases, or disorders, T-cell and B-cell neoplasms,
CC inflammation disorders, diseases and conditions, rheumatoid arthritis,
CC osteoarthritis, psoriasis, rhinitis, inflammatory bowel disease (Crohn's
CC and ulcerative colitis), allergies, particularly those involving
CC hyperactivity of B-cells and T- cells, or other immune cells, such as
CC mast cells or eosinophils, autoimmune diseases such as systemic lupus
CC erythematosus and multiple sclerosis, pulmonary diseases including
CC asthma, acute respiratory distress syndrome, and chronic obstructive
CC pulmonary disorder, tissue/ organ rejection and cancer. The invention is
CC useful in gene therapy. The present sequence is human SLAP-2 cDNA
XX
SQ Sequence 2567 BP; 611 A; 741 C; 666 G; 549 T; 0 U; 0 Other;

```
  Query Match             100.0%;  Score 786;  DB 6;  Length 2567;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-197;
  Matches  786;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        415 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 474

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        475 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 534
```

```
Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        535 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 594

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        595 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 654

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        655 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 714

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        715 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 774

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        775 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 834

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        835 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 894

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        895 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 954

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        955 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 1014

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1015 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 1074

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1075 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 1134

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1135 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 1194

Qy        781 GCCTAG 786
              ||||||
Db       1195 GCCTAG 1200
```

RESULT 4
ADF90741/c
ID ADF90741 standard; DNA; 3646 BP.
XX
AC ADF90741;
XX
DT 26-FEB-2004 (first entry)
XX
DE Human hepatic-fibrosis disease marker SEQ ID 203.
XX
KW Hepatic fibrosis; marker; chronic hepatitis; liver cirrhosis;
KW hepatic carcinoma; human; ds.
XX
OS Homo sapiens.
XX
PN JP2003259877-A.
XX
PD 16-SEP-2003.
XX
PF 11-MAR-2002; 2002JP-00065013.
XX
PR 11-MAR-2002; 2002JP-00065013.
XX
PA (SUMU) SUMITOMO SEIYAKU KK.
XX
DR WPI; 2003-821598/77.
XX
PT Hepatic fibrosis disease markers comprising polynucleotides or
PT antibodies, useful for improved diagnosis, screening and developing drugs
PT to treat hepatitis, to control cirrhosis and carcinoma.
XX
PS Claim 1; SEQ ID NO 203; 313pp; Japanese.
XX
CC The present invention relates to hepatic-fibrosis disease markers
CC (ADF90539-ADF90871) and related proteins (ADF90872-ADF90917). The
CC sequences are useful for detecting and treating hepatic fibrosis caused
CC by alcohol consumption, virus infection, etc., and the associated chronic
CC hepatitis, etc. leading to liver cirrhosis and hepatic carcinoma. The
CC markers allow the cause of hepatic fibrosis to be clarified (diagnostic
CC precision), so more suitable treatments can be developed and given.
XX
SQ Sequence 3646 BP; 782 A; 954 C; 1031 G; 877 T; 0 U; 2 Other;

Query Match 100.0%; Score 786; DB 10; Length 3646;
Best Local Similarity 100.0%; Pred. No. 1.7e-197;
Matches 786; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3202 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 3143

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3142 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 3083

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3082 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 3023

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3022 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 2963

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2962 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 2903

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2902 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 2843

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2842 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 2783

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2782 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 2723

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2722 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 2663

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2662 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 2603

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2602 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 2543

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2542 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 2483

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2482 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 2423

Qy        781 GCCTAG 786
              ||||||
Db       2422 GCCTAG 2417
```

RESULT 5
ADF82458
ID ADF82458 standard; DNA; 2788 BP.
XX
AC ADF82458;
XX
DT 26-FEB-2004 (first entry)
XX
DE Leukaemia-related DNA sequence #3014.
XX
KW Cytostatic; Gene therapy; leukaemia; ss.
XX
OS Unidentified.
XX
PN WO2003039443-A2.
XX
PD 15-MAY-2003.
XX
PF 04-NOV-2002; 2002WO-EP012303.
XX
PR 05-NOV-2001; 2001EP-00126244.
PR 30-APR-2002; 2002EP-00009758.
XX
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA (UYLU-) UNIV LUDWIG MAXIMILIANS.
PA (HAFE/) HAFERLACH T.
PA (SCHO/) SCHOCH C.
PA (KERN/) KERN W.
XX
PI Haferlach T, Schoch C, Kern W, Kohlmann A, Schnittger S, Dugas M;
PI Eils R, Brors B, Mergenthaler S;
XX
DR WPI; 2003-505037/47.
XX
PT Determining the subtype of leukemia cells and whether a patient sample
PT contains leukemia cells or other cells, useful for treating leukemia,
PT comprises determining the expression profile of a group of markers in a
PT patient sample.
XX
PS Disclosure; SEQ ID NO 3014; 2938pp; English.
XX
CC The present invention relates to a method (M1) for determining the
CC subtype of leukaemia cells and whether a patient sample contains
CC leukaemia cells. The method comprises determining the expression profile
CC of a group of markers in a patient sample. The method is useful for
CC determining the presence of leukaemia cells, its types or subtypes, and
CC for the preparation of a medicament for treating leukaemia.
XX
SQ Sequence 2788 BP; 663 A; 764 C; 700 G; 595 T; 0 U; 66 Other;

Query Match 99.9%; Score 785; DB 10; Length 2788;
Best Local Similarity 99.9%; Pred. No. 2.8e-197;
Matches 785; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 446

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 CAAGGCCANGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 506

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 566

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        567 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 626

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        627 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 686

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        687 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 746

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        747 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 806

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        807 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 866

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        867 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 926

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        927 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 986

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        987 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 1046

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1047 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 1106

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1107 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 1166

Qy        781 GCCTAG 786
              ||||||
Db       1167 GCCTAG 1172
```

RESULT 6
ABK61465
ID ABK61465 standard; cDNA; 1183 BP.
XX
AC ABK61465;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human cDNA encoding protein NOV13.
XX
KW Human; gene; ss; NOVX; gene therapy; cardiomyopathy; atherosclerosis;
KW cell signal processing disorder; metabolic pathway modulation disorder;
KW diabetes; cancer; adenocarcinoma; lymphoma; prostate cancer;
KW uterus cancer; immune response; graft-versus-host disease;
KW acquired immunodeficiency syndrome; AIDS; asthma; Crohn's disease;
KW hypertension; congenital heart defects; multiple sclerosis; inflammation;
KW Albright hereditary osteodystrophy.
XX
OS Homo sapiens.
XX
PN WO200216599-A2.
XX
PD 28-FEB-2002.
XX
PF 27-AUG-2001; 2001WO-US026510.
XX
PR 25-AUG-2000; 2000US-0228191P.
PR 08-FEB-2001; 2001US-0267300P.

PR 20-FEB-2001; 2001US-0269961P.
PR 20-MAR-2001; 2001US-0277337P.
XX
PA (CURA-) CURAGEN CORP.
PA (CORT-) COR THERAPEUTICS INC.
XX
PI Burgess CE, Conley PB, Grosse WM, Hart M, Kekuda R, Shimkets RA;
PI Spytek KA, Szekeres ES, Tomlinson JE, Topper JN, Yang R;
XX
DR WPI; 2002-280937/32.
DR P-PSDB; AAU91308.
XX
PT New polypeptides for treating or preventing a disorder associated with
PT them, in humans, e.g. cardiomyopathy, atherosclerosis or cancers.
XX
PS Claim 1; Page 98; 263pp; English.
XX
CC The invention relates to an isolated polypeptide (NOVX) a mature form of
CC NOVX, a NOVX variant (differing by no more than 15%), the nucleotide
CC encoding NOVX (or its complement, fragment or variant). NOVX is NOV1-14,
CC 15a, 15b, 16a, and 16b. The NOVX polypeptide, nucleic acid encoding it
CC and antibody against it, are useful for treating or preventing (e.g. by
CC gene therapy) a NOVX-associated disorder in humans, e.g. cardiomyopathy,
CC atherosclerosis, a disorder related to cell signal processing and
CC metabolic pathway modulation, diabetes or cancers. The NOVX polypeptide
CC and nucleic acids are also useful for determining the presence of
CC predisposition to the diseases. The NOVX nucleic acid and polypeptide are
CC especially useful in therapeutic or prophylactic applications for
CC disorders associated with aberrant NOVX expression or activity, e.g.
CC cancers (e.g. adenocarcinoma, lymphoma, prostate cancer or uterus
CC cancer), immune response, graft-versus-host disease, acquired
CC immunodeficiency syndrome (AIDS), asthma, Crohn's disease, hypertension,
CC congenital heart defects, multiple sclerosis, inflammation or Albright
CC hereditary osteodystrophy and many other diseases listed in the
CC specification. The DNA encoding the protein is useful in gene therapy for
CC treating the conditions. This is also useful in detection assays,
CC chromosome mapping, tissue typing, diagnostic or prognostic assays, or
CC for developing a powerful assay system for functional analysis of various
CC human disorders, as well as in diagnostic applications. The present
CC sequence encodes a NOVX protein
XX
SQ Sequence 1183 BP; 251 A; 359 C; 333 G; 240 T; 0 U; 0 Other;

Query Match 99.8%; Score 784.4; DB 6; Length 1183;
Best Local Similarity 99.9%; Pred. No. 3.2e-197;
Matches 785; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        398 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 457

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        458 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 517

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        518 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 577

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        578 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 637

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        638 AACATCCCCAGCGTCCACGTGGGCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 697

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        698 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 757

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        758 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 817

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        818 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 877

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        878 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 937

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        938 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 997

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        998 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 1057

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1058 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 1117

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1118 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 1177

Qy        781 GCCTAG 786
              ||||||
Db       1178 GCCTAG 1183
```

RESULT 7
AAC77202
ID AAC77202 standard; cDNA; 837 BP.
XX
AC AAC77202;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2757 polynucleotide sequence SEQ ID NO:5513.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US008621.
XX
PR 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX

DR WPI; 2000-602362/57.
DR P-PSDB; AAB42993.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
PS Claim 5; Page 4692-4693; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ Sequence 837 BP; 176 A; 254 C; 245 G; 160 T; 0 U; 2 Other;

Query Match 94.4%; Score 741.8; DB 3; Length 837;
Best Local Similarity 99.7%; Pred. No. 5.2e-186;
Matches 743; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy        42 AAGCTTGAGTTCCTCTGTCCAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAA 101
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3 AAGCTTGAGTTCCTCTGTCCAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAA 62

Qy       102 GGCCACAGCCGTGGCCCTGGGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAG 161
             |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63 GGCCACAGCCGCGGCCCTGGGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAG 122

Qy       162 ACTCGGGGAGCCATTGACCATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGA 221
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       123 ACTCGGGGAGCCATTGACCATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGA 182

Qy       222 AGTCTCAGGCAGAGAGTATAACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTG 281
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       183 AGTCTCAGGCAGAGAGTATAACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTG 242

Qy       282 GCTGTATGAGGGCCTGAGCAGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCC 341
             |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       243 GCTGTATGAGGGCCTGAGCAGGGAGAAAGCAGAGGACCTGCTGTTGTTACCTGGGAACCC 302

Qy       342 TGGAGGGGCCTTCCTCATCCGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGT 401
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       303 TGGAGGGGCCTTCCTCATCCGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGT 362

Qy       402 CCGCCTCAGCCGCCCTGCATCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGA 461
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       363 CCGCCTCAGCCGCCCTGCATCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGA 422

Qy       462 CAATGGCTGGCTGTACATCTCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGA 521
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       423 CAATGGCTGGCTGTACATCTCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGA 482

Qy       522 CCATTACTCTGAGCTGGCGGATGACATCTGCTGCCTACTCAAGGAGCCCCTGTGTCCTGCA 581
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       483 CCATTACTCTGAGCTGGCGGATGACATCTGCTGCCTACTCAAGGAGCCCCTGTGTCCTGCA 542

Qy       582 GAGGGCTGGCCCGCTCCCTGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACC 641
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       543 GAGGGCTGGCCCGCTCCCTGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACC 602

Qy       642 ACTCAACTGGAAAGAGCTGGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGA 701
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       603 ACTCAACTGGAAAGAGCTGGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGA 662

Qy       702 GTCTCTTCTCAGTGAGGGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGA 761
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       663 GTCTCTTCTCAGTGAGGGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGA 722

Qy       762 GGCTGTCTCTTTGGATGATGCCTAG 786
             |||||||||||||||||||||||||
Db       723 GGCTGTCTCTTTGGATGATGCCTAG 747
```

RESULT 8
ABQ99374
ID ABQ99374 standard; cDNA; 1413 BP.
XX
AC ABQ99374;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human coding sequence SEQ ID 107.
XX
KW Human; expressed sequence tag; EST; chromosome 20;
KW haematopoietic disorder; central nervous system disease; viral infection;
KW peripheral nervous system disease; non-healing wound; infectious disease;
KW immune deficiency; immune disorder; bacterial infection; allergy; cancer;
KW fungal infection; autoimmune disorder; coagulation disorder; nootropic;
KW antiallergic; antiinflammatory; immunosuppressive; neuroprotective;
KW cytostatic; haemostatic; virucide; antibacterial; fungicide;
KW immunostimulant; cerebroprotective; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200259260-A2.
XX
PD 01-AUG-2002.
XX
PF 16-NOV-2001; 2001WO-US042950.
XX
PR 17-NOV-2000; 2000US-00714936.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;
PI Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR WPI; 2002-590824/63.
DR N-PSDB; ABP64788.
XX
PT New isolated polynucleotide, useful in research, diagnostic or
PT therapeutic methods, e.g. preventing or treating disorders involving
PT aberrant protein expression or biological activity.
XX
PS Claim 1; SEQ ID NO 107; 394pp; English.
XX
CC The present invention relates to novel human coding sequences (ABQ99268-
CC ABQ99608) and proteins (ABP64682-ABP65022). The sequences are useful in
CC therapeutic, diagnostic and research methods. The polynucleotides may be
CC used in the field of molecular biology as hybridisation probes, primers
CC for PCR, for chromosome and gene mapping, for the recombinant production
CC of protein, or in generation of anti-sense DNA or RNA. The
CC polynucleotides are useful in diagnostics as expressed sequence tags
CC (ESTs) for identifying expressed genes or for physical mapping of the
CC human genome. The proteins may be used as molecular weight markers, or as
CC nutritional sources or supplements. The proteins may be used to maintain
CC and expand cell population in a totipotential or pluripotential state
CC useful for re-engineering damaged or diseased tissues, transplantation,
CC manufacture of bio-pharmaceuticals or the development of bio-sensors. The
CC polynucleotides and proteins are useful for preventing, treating or

```
CC   ameliorating disorders involving aberrant protein expression or
CC   biological activity, e.g. haematopoietic disorders, central/peripheral
CC   nervous system diseases, mechanical and traumatic disorders, non-healing
CC   wounds, immune deficiencies and disorders, infectious diseases caused by
CC   viral, bacterial or fungal infection, autoimmune disorders, allergic
CC   reactions and conditions, coagulation disorders, or cancer. The
CC   polynucleotide sequences of the invention were assembled from ESTs
CC   isolated mainly by sequencing by hybridisation, and in some cases,
CC   sequences obtained from one or more public databases. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1413 BP; 332 A; 407 C; 388 G; 286 T; 0 U; 0 Other;

  Query Match             85.6%;  Score 672.8;  DB 6;  Length 1413;
  Best Local Similarity   93.4%;  Pred. No. 1.1e-167;
  Matches  734;  Conservative    0;  Mismatches    2;  Indels   50;  Gaps    1;

Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         54 ATGGGAGGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 113

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        114 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 173

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        174 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 233

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        234 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 293

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        294 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 353

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        354 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 413

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        414 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 473

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        474 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 533

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              |||||||||||||||||||||||||||||||||| ||||||||||||
Db        534 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGGGGGACCATTACTCT--------- 584

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
                                                  ||||||||||||||||||||
Db        585 ----------------------------------------GAGGGCTGGCCCGCTCCCT 603

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 663

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 723

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 783

Qy        781 GCCTAG 786
              ||||||
Db        784 GCCTAG 789


RESULT 9
AAL44090
ID   AAL44090 standard; cDNA; 737 BP.
XX
AC   AAL44090;
XX
DT   03-OCT-2002  (first entry)
XX
DE   Mouse MARS short isoform protein coding sequence.
XX
KW   Mouse; gene; ss; gene therapy; modulator of antigen receptor signalling;
KW   MARS; tumour suppressor gene; Scr-like adaptor protein; SLAP;
KW   myeloid malignancy; acute myelogenous leukaemia; autoimmune disorder;
KW   immunosuppression; myeloproliferative disorder; breast cancer.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .633
FT                   /*tag=  a
FT                   /product= "Mouse MARS short isoform protein"
XX
PN   WO200242452-A2.
XX
PD   30-MAY-2002.
XX
PF   26-NOV-2001; 2001WO-CA001662.
XX
PR   27-NOV-2000; 2000CA-02324663.
XX
PA   (HOSP-) HOSPITAL FOR SICK CHILDREN.
XX
PI   Mcglade JC,  Loreto MP;
XX
DR   WPI; 2002-566564/60.
DR   P-PSDB; AAO15458.
XX
PT   New isolated modulator of antigen receptor signaling protein or its
PT   fragment, useful for treating malignant disorders such as myeloid
PT   malignancies, autoimmune disorders and myeloproliferative disorders.
XX
PS   Claim 9; Page 77; 110pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of modulator
CC   of antigen receptor signalling (MARS) proteins. The MARS protein is a
CC   putative tumour suppressor gene and exhibits structural and sequence
CC   similarity to the Scr-like adaptor protein (SLAP). The MARS DNA and
CC   protein sequences of the invention are useful for the treatment of
CC   myeloid malignancies (e.g. acute myelogenous leukaemia) autoimmune
CC   disorders, immunosuppression, myeloproliferative disorders and
CC   malignancies related to the de-regulation of tyrosine kinases (e.g.
CC   breast cancer). The present cDNA sequence encodes a mouse MARS protein
XX
SQ   Sequence 737 BP; 152 A; 219 C; 218 G; 148 T; 0 U; 0 Other;

  Query Match             84.4%;  Score 663.4;  DB 6;  Length 737;
  Best Local Similarity   93.4%;  Pred. No. 2.7e-165;
  Matches  735;  Conservative    0;  Mismatches    1;  Indels   51;  Gaps    2;

Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
```

```
Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCT--------- 531

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
                                                  ||||||||||||||||||||
Db        532 ------------------------------------------GAGGGCTGGCCCGCTCCCT 550

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        551 GGCAAGGATATACCCCTACCTGTGACTGTGCGGAGGACACCACTCAACTGGAAAGAGCTG 610

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        611 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 670

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGA-CGAGGCTGTCTCTTTGGATGA 779
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        671 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGAGCGAGGCTGTCTCTTTGGATGA 730

Qy        780 TGCCTAG 786
              |||||||
Db        731 TGCCTAG 737
```

RESULT 10
AAS74750
ID AAS74750 standard; cDNA; 2049 BP.
XX
AC AAS74750;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #10554.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG10563.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 10554; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2049 BP; 479 A; 573 C; 551 G; 443 T; 0 U; 3 Other;

Query Match 77.6%; Score 609.8; DB 5; Length 2049;
Best Local Similarity 92.2%; Pred. No. 5.6e-151;
Matches 718; Conservative 0; Mismatches 7; Indels 54; Gaps 5;

```
Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        965 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 1024

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1025 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 1084

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1085 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 1144

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1145 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 1204

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1205 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 1264

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1265 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 1324

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       1325 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAACCGCCCTGCA 1384

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              | |||||||||||||||||||||||||||| ||||| ||||||||||||||||||||||
Db       1385 TTCTGGGACCGGATCAGACACTACAGGATTCACTGGCTTGACAATGGCTGGCTGTACATC 1444

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1445 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCT--------- 1495

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
                                                  ||||||||||||||||||||
Db       1496 ----------------------------------------GAGGGCTGGCCCGCTCCCT 1514

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1515 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 1574

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACA-GGGGAGGAGTCTCTTCTCAG-TGAGG 718
               |||||||||||||||||||||||||||||||| |||||||||||||||||||   ||||
Db       1575 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGGAGGAGTCTCTTCTCAGAGGAGG 1634

Qy        719 GTCTCCGGGAGT-CCCTCAGCTTCTACATCAG-CCTGAATGACGAGGCTGTCTCTTTGG 775
               | ||||||||||| |||||||||||||||| |||||||||||||||||||||||| |
Db       1635 GGCTCCGGGAGTCCCCTCAGCTTCTACATCAGCCCTGAATGACGAGGCTGTCTCTTTTG 1693
```

RESULT 11
AAL44087
ID AAL44087 standard; cDNA; 1348 BP.
XX
AC AAL44087;
XX
DT 03-OCT-2002 (first entry)
XX
DE Mouse modulator of antigen receptor signalling protein coding sequence.
XX
KW Mouse; gene; ss; gene therapy; modulator of antigen receptor signalling;
KW MARS; tumour suppressor gene; Scr-like adaptor protein; SLAP;
KW myeloid malignancy; acute myelogenous leukaemia; autoimmune disorder;
KW immunosuppression; myeloproliferative disorder; breast cancer.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 282. .1061
FT /*tag= a
FT /product= "Mouse MARS protein"
XX
PN WO200242452-A2.
XX
PD 30-MAY-2002.
XX
PF 26-NOV-2001; 2001WO-CA001662.
XX
PR 27-NOV-2000; 2000CA-02324663.
XX
PA (HOSP-) HOSPITAL FOR SICK CHILDREN.
XX
PI Mcglade JC, Loreto MP;
XX
DR WPI; 2002-566564/60.
DR P-PSDB; AAO15456.
XX
PT New isolated modulator of antigen receptor signaling protein or its
PT fragment, useful for treating malignant disorders such as myeloid
PT malignancies, autoimmune disorders and myeloproliferative disorders.
XX
PS Claim 10; Fig 1A; 110pp; English.
XX
CC The invention comprises the amino acid and coding sequences of modulator
CC of antigen receptor signalling (MARS) proteins. The MARS protein is a
CC putative tumour suppressor gene and exhibits structural and sequence
CC similarity to the Scr-like adaptor protein (SLAP). The MARS DNA and
CC protein sequences of the invention are useful for the treatment of
CC myeloid malignancies (e.g. acute myelogenous leukaemia) autoimmune
CC disorders, immunosuppression, myeloproliferative disorders and
CC malignancies related to the de-regulation of tyrosine kinases (e.g.
CC breast cancer). The present cDNA sequence encodes a mouse MARS protein
XX
SQ Sequence 1348 BP; 324 A; 385 C; 362 G; 277 T; 0 U; 0 Other;

Query Match 68.4%; Score 537.8; DB 6; Length 1348;
Best Local Similarity 81.9%; Pred. No. 5.4e-132;
Matches 646; Conservative 0; Mismatches 137; Indels 6; Gaps 2;

```
Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||| || |||||||| | ||| |     | ||||| |||    || |||||||
Db        282 ATGGGAAGTTTGTCCAGCAGAGGGAAAAC---CTCCAGCCCCAGCCCCAGCTCCTCTGGT 338

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              | || ||||| ||| ||| ||||| || |||| ||| |||||  |||||| |||||||||
Db        339 CCAGACCAGGAACCCGTGTCCATGCAACCAGAAAGACACAAGGTCACAGCTGTGGCCCTG 398

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              |||||||||| |||||||  | ||||   || || ||||||||||||||||  |||||
Db        399 GGCAGTTTCCCAGCAGGTGAACAGGCCAGACTATCTCTGAGACTCGGGGAGCCGCTGACC 458

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||| |||||||||||||||| |||||||| || | ||| ||||||||||||||||||||
Db        459 ATCATCTCTGAGGATGGAGATTGGTGGACAGTCCAGTCGGAAGTCTCAGGCAGAGAGTAC 518

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
                |||| ||||| ||  | ||||| |||||| |||| |||||||||| ||||||||||
Db        519 CACATGCCCAGTGTGTATGTGGCTAAAGTCGCCCACGGGTGGCTGTACGAGGGCCTGAGC 578

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              |||||||||| |||||||| ||  |||||||||||||||| ||||||||||||||||||
Db        579 CGGGAGAAAGCCGAGGAACTACTCCTGTTACCTGGGAACCCCGGAGGGGCCTTCCTCATC 638

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              |||||||||||||||||||||||||  || || ||||| ||||| ||||||||||||||
Db        639 CGGGAGAGCCAGACCAGGAGAGGCTGCTATTCCCTGTCCGTCCGACTCAGCCGCCCTGCA 698

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              || |||||||||||||||||||||||||||||| || | ||||||||||||||||||||
Db        699 TCTTGGGACCGGATCAGACACTACAGGATACAGCGTCTTGACAATGGCTGGCTGTACATC 758

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||| |||||||||||||||||||||| ||| ||||||| |||||||||||||| ||
Db        759 TCACCTCGCCTCACCTTCCCCTCACTCCACGCCTTGGTGGAGCATTACTCTGAGCTAGCA 818

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              |||| ||||||||| |  |||| |||||| |||||||||||| |  ||| || || |||
Db        819 GATGGCATCTGCTGTCCCCTCAGGGAGCCGTGTGTCCTGCAGAAGCTTGGGCCACTACCT 878

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||| |||| ||| | |||||||||||||||  |   || |||| || |||||| |||||
Db        879 GGCAAAGATACACCTCCACCTGTGACTGTGCCAACATCATCACTAAATTGGAAAAAGCTG 938

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAG---CTGCCACAGGGGAGGAGTCTCTTCTCAGTGAG 717
              ||| ||  |||||||||  ||||  |||| |  |||||||  ||||| |||||||||
Db        939 GACCGCAGCCTCCTGTTTCTGGAAGCACCTGCGAGTGGGGAGGCATCTCTGCTCAGTGAG 998

Qy        718 GGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGAT 777
              || ||||| |||||||||||| | ||||||||||||||  ||| ||  |  |   | | |
Db        999 GGGCTCCGAGAGTCCCTCAGTTCCTACATCAGCCTGGCTGAGGACCCCTTGGATGATGCT 1058

Qy        778 GATGCCTAG 786
                |  ||| |
Db       1059 TAGCCCTGG 1067
```

RESULT 12
ABQ98670
ID ABQ98670 standard; DNA; 763 BP.
XX
AC ABQ98670;
XX
DT 04-NOV-2002 (first entry)
XX
DE Human ORF477 coding sequence.
XX
KW Cytostatic; Cardiant; Anti-allergic; Immunosuppressive; Vulnerary;

KW Antiinflammatory; gene therapy; human; ORFX; atherogenic; platelet;
KW human umbilical vein endothelial cell; HUVEC; atherosclerotic plaque;
KW cancer; cardiovascular disease; allergy; autoimmune disease;
KW wound healing; blood coagulation disorder; inflammatory disorder; ds.
XX
OS Homo sapiens.
XX
PN US2002082206-A1.
XX
PD 27-JUN-2002.
XX
PF 30-MAY-2001; 2001US-00867550.
XX
PR 30-MAY-2000; 2000US-0208427P.
XX
PA (LEAC/) LEACH M D.
PA (MEHR/) MEHRABAN F.
PA (CONL/) CONLEY P B.
PA (TOPP/) TOPPER J N.
PA (LAWD/) LAW D.
XX
PI Leach MD, Mehraban F, Conley PB, Topper JN, Law D;
XX
DR WPI; 2002-626554/67.
DR P-PSDB; ABP64107.
XX
PT New polypeptide designated ORFX are present in human atherogenic cells
PT and are useful to prevent and treat ORFX-associated disorders including
PT cancer, allergy, wound healing or autoimmune, cardiovascular or
PT inflammatory disease.
XX
PS Claim 2; SEQ ID NO 953; 78pp; English.
XX
CC The present invention relates to novel human ORFX polypeptides and their
CC coding sequences (ABP63631-ABP64681 and ABQ98194-ABQ99267). The sequences
CC were discovered in human atherogenic cells, in particular in platelets
CC and human umbilical vein endothelial cells (HUVEC) and are expressed in
CC many other tissues as well. Atherogenic cells are cells which have the
CC potential to develop atherosclerotic plaques. The ORFX polypeptides and
CC nucleic acids are useful for treating or preventing a pathological
CC condition associated with an ORFX-associated disorder, e.g. cancer,
CC cardiovascular disease, allergy, autoimmune disease, wound healing, blood
CC coagulation disorders or inflammatory disorders. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from the USPTO web site at
CC seqdata.uspto.gov/sequence.html?DocID=20020082206
XX
SQ Sequence 763 BP; 176 A; 222 C; 218 G; 147 T; 0 U; 0 Other;

Query Match 60.8%; Score 478; DB 6; Length 763;
Best Local Similarity 100.0%; Pred. No. 3e-116;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        286 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 345

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        346 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 405

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        406 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 465

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        466 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 525

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        526 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 585

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        586 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 645

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        646 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 705

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACA 478
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        706 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACA 763
```

RESULT 13
AAS74748
ID AAS74748 standard; cDNA; 603 BP.
XX
AC AAS74748;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #10552.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG10561.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 10552; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

```
SQ   Sequence 603 BP; 124 A; 189 C; 164 G; 126 T; 0 U; 0 Other;

  Query Match             51.5%;  Score 405;  DB 5;  Length 603;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-97;
  Matches  405;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        382 GGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCATCCTGGGACCGGATCAGACAC 441
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 GGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCATCCTGGGACCGGATCAGACAC 258

Qy        442 TACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATCTCACCGCGCCTCACCTTCCCC 501
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 TACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATCTCACCGCGCCTCACCTTCCCC 318

Qy        502 TCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCGGATGACATCTGCTGCCTACTC 561
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 TCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCGGATGACATCTGCTGCCTACTC 378

Qy        562 AAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCTGGCAAGGATATACCCCTACCT 621
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 AAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCTGGCAAGGATATACCCCTACCT 438

Qy        622 GTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTGGACAGCTCCCTCCTGTTTTCT 681
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 GTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTGGACAGCTCCCTCCTGTTTTCT 498

Qy        682 GAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGTCTCCGGGAGTCCCTCAGCTTC 741
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 GAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGTCTCCGGGAGTCCCTCAGCTTC 558

Qy        742 TACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGATGCCTAG 786
              |||||||||||||||||||||||||||||||||||||||||||||
Db        559 TACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGATGCCTAG 603
```

```
RESULT 14
ADL63090
ID   ADL63090 standard; DNA; 864 BP.
XX
AC   ADL63090;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Human ovarian cancer DNA marker #21302.
XX
KW   Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
XX
OS   Homo sapiens.
XX
PN   WO200170979-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US009126.
XX
PR   21-MAR-2000; 2000US-0191031P.
PR   25-MAY-2000; 2000US-0207124P.
PR   15-JUN-2000; 2000US-0211940P.
PR   07-JUL-2000; 2000US-0216820P.
PR   25-JUL-2000; 2000US-0220661P.
PR   21-DEC-2000; 2000US-0257672P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lee J,  Lillie J;
XX
DR   WPI; 2001-611502/70.
XX
PT   Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
PT   cancer cells as compared to their normal non-cancerous ovarian cells are
PT   used to characterize stage, grade, histological type of ovarian cancer.
XX
PS   Disclosure; SEQ ID NO 21302; 106pp; English.
XX
CC   The invention relates to nucleic acid markers which are overexpressed in
CC   ovarian cancer cells as compared to their expression in normal (i.e. non-
CC   cancerous) ovarian cells. The invention also relates to polypeptides
CC   encoded by the markers, antibodies that selectively bind to the
CC   polypeptides, a method of inhibiting ovarian cancer in a patient at risk
CC   of developing ovarian cancer involving inhibiting expression of a gene
CC   corresponding to a marker of the invention and a method of treating a
CC   patient afflicted with ovarian cancer comprising providing to cells of
CC   the patient an antisense oligonucleotide complementary to a marker of the
CC   invention. The markers are useful for assessing if a patient is afflicted
CC   with ovarian cancer, which involves comparing the level of expression of
CC   a marker in a patient sample and a normal level of expression of the
CC   marker in a control non-ovarian cancer sample. A difference between the
CC   expression levels indicates ovarian cancer. The level of expression of a
CC   marker corresponds to a secreted protein or to a transcribed
CC   polynucleotide or its portion. The level of expression of the marker is
CC   assessed by detecting the presence in the sample, a protein or protein
CC   fragment corresponding to the marker. The presence of protein or protein
CC   fragment is detected using an antibody that specifically binds with the
CC   protein or protein fragment. Alternatively, the level of expression of
CC   the marker is assessed by detecting the presence of a transcribed
CC   polynucleotide which anneals with the marker or anneals with a portion of
CC   the polynucleotide comprising the marker, under stringent conditions. The
CC   marker is also used for monitoring the progression of ovarian cancer in a
CC   patient which involves detecting expression of the marker in a patient
CC   sample at a first point in time, repeating the method at a subsequent
CC   time and comparing the level of expression. The method is carried out
CC   using an ovarian tissue sample. A composition comprising a marker,
CC   polypeptide or antibody of the invention is used to treat ovarian cancer.
CC   This sequence represents a human ovarian cancer DNA marker of the
CC   invention.
XX
SQ   Sequence 864 BP; 191 A; 259 C; 242 G; 165 T; 0 U; 7 Other;

  Query Match             48.5%;  Score 381;  DB 5;  Length 864;
  Best Local Similarity   98.7%;  Pred. No. 1.4e-90;
  Matches  384;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        450 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 509

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        510 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 569

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        570 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 629

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        630 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 689

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 749

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        750 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 809

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTA 389
              |||||||||||||||||||    | || ||
Db        810 CGGGAGAGCCAGACCAGGAAGAGGTCCTA 838
```

```
RESULT 15
ABQ99151
ID   ABQ99151 standard; DNA; 875 BP.
XX
AC   ABQ99151;
```

XX
DT 04-NOV-2002 (first entry)
XX
DE Human ORF958 coding sequence.
XX
KW Cytostatic; Cardiant; Anti-allergic; Immunosuppressive; Vulnerary;
KW Antiinflammatory; gene therapy; human; ORFX; atherogenic; platelet;
KW human umbilical vein endothelial cell; HUVEC; atherosclerotic plaque;
KW cancer; cardiovascular disease; allergy; autoimmune disease;
KW wound healing; blood coagulation disorder; inflammatory disorder; ds.
XX
OS Homo sapiens.
XX
PN US2002082206-A1.
XX
PD 27-JUN-2002.
XX
PF 30-MAY-2001; 2001US-00867550.
XX
PR 30-MAY-2000; 2000US-0208427P.
XX
PA (LEAC/) LEACH M D.
PA (MEHR/) MEHRABAN F.
PA (CONL/) CONLEY P B.
PA (TOPP/) TOPPER J N.
PA (LAWD/) LAW D.
XX
PI Leach MD, Mehraban F, Conley PB, Topper JN, Law D;
XX
DR WPI; 2002-626554/67.
DR P-PSDB; ABP64588.
XX
PT New polypeptide designated ORFX are present in human atherogenic cells
PT and are useful to prevent and treat ORFX-associated disorders including
PT cancer, allergy, wound healing or autoimmune, cardiovascular or
PT inflammatory disease.
XX
PS Claim 2; SEQ ID NO 1915; 78pp; English.
XX
CC The present invention relates to novel human ORFX polypeptides and their
CC coding sequences (ABP63631-ABP64681 and ABQ98194-ABQ99267). The sequences
CC were discovered in human atherogenic cells, in particular in platelets
CC and human umbilical vein endothelial cells (HUVEC) and are expressed in
CC many other tissues as well. Atherogenic cells are cells which have the
CC potential to develop atherosclerotic plaques. The ORFX polypeptides and
CC nucleic acids are useful for treating or preventing a pathological
CC condition associated with an ORFX-associated disorder, e.g. cancer,
CC cardiovascular disease, allergy, autoimmune disease, wound healing, blood
CC coagulation disorders or inflammatory disorders. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from the USPTO web site at
CC seqdata.uspto.gov/sequence.html?DocID=20020082206
XX
SQ Sequence 875 BP; 205 A; 259 C; 225 G; 185 T; 0 U; 1 Other;

Query Match 43.4%; Score 341; DB 6; Length 875;
Best Local Similarity 100.0%; Pred. No. 5.3e-80;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        446 GGATCCACTGCCTTGACAATGGCTGGCTGTACATCTCACCGCGCCTCACCTTCCCCTCAC 505
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GGATCCACTGCCTTGACAATGGCTGGCTGTACATCTCACCGCGCCTCACCTTCCCCTCAC 61

Qy        506 TCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCGGATGACATCTGCTGCCTACTCAAGG 565
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 TCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCGGATGACATCTGCTGCCTACTCAAGG 121

Qy        566 AGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCTGGCAAGGATATACCCCTACCTGTGA 625
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 AGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCTGGCAAGGATATACCCCTACCTGTGA 181

Qy        626 CTGTGCAGAGGACACCACTCAACTGGAAAGAGCTGGACAGCTCCCTCCTGTTTTCTGAAG 685
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 CTGTGCAGAGGACACCACTCAACTGGAAAGAGCTGGACAGCTCCCTCCTGTTTTCTGAAG 241

Qy        686 CTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGTCTCCGGGAGTCCCTCAGCTTCTACA 745
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 CTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGTCTCCGGGAGTCCCTCAGCTTCTACA 301

Qy        746 TCAGCCTGAATGACGAGGCTGTCTCTTTGGATGATGCCTAG 786
              ||||||||||||||||||||||||||||||||||||||||
Db        302 TCAGCCTGAATGACGAGGCTGTCTCTTTGGATGATGCCTAG 342
```

Search completed: November 16, 2004, 21:33:29
Job time : 518 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 16, 2004, 17:07:09 ; Search time 3766 Seconds
(without alignments)
9869.810 Million cell updates/sec

Title: US-10-043-649-1
Perfect score: 786
Sequence: 1 atgggaagtctgcccagcag.....tctctttggatgatgcctag 786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	786	100.0	786	6	AX511153	AX511153 Sequence
	2	786	100.0	786	6	AX572845	AX572845 Sequence
	3	786	100.0	786	9	AF290985	AF290985 Homo sapi
	4	786	100.0	786	9	AF326353	AF326353 Homo sapi
	5	786	100.0	2415	9	AK025645	AK025645 Homo sapi
	6	786	100.0	2538	9	BC042041	BC042041 Homo sapi
	7	786	100.0	2567	6	AX452880	AX452880 Sequence
	8	785	99.9	2788	6	AX780857	AX780857 Sequence
	9	784.4	99.8	1183	6	AX443133	AX443133 Sequence
c	10	784.4	99.8	1183	6	AX443135	AX443135 Sequence
	11	663.4	84.4	737	6	AX511155	AX511155 Sequence
	12	663.4	84.4	737	9	AF290986	AF290986 Homo sapi
	13	539.4	68.6	1384	10	BC052655	BC052655 Mus muscu
	14	537.8	68.4	1348	6	AX511150	AX511150 Sequence
	15	537.8	68.4	1348	10	AF287467	AF287467 Mus muscu
	16	536.8	68.3	777	6	AX511151	AX511151 Sequence
	17	536.2	68.2	1321	10	AF434990	AF434990 Mus muscu
	18	441	56.1	441	6	CQ724864	CQ724864 Sequence
	19	381	48.5	864	6	CQ414231	CQ414231 Sequence
	20	215	27.4	320	6	CQ410243	CQ410243 Sequence
	21	214.2	27.3	1571	5	BC056035	BC056035 Xenopus l
	22	160.6	20.4	831	9	CR536537	CR536537 Homo sapi
	23	160.6	20.4	2021	9	HSU44403	U44403 Human Src-l
	24	160.6	20.4	2670	6	CQ725810	CQ725810 Sequence
	25	160.6	20.4	3090	9	HSM808530	BX648382 Homo sapi
	26	160.2	20.4	2109	6	AX428893	AX428893 Sequence
	27	159	20.2	1076	9	HSU30473	U30473 Homo sapien
	28	159	20.2	1870	9	BC007042	BC007042 Homo sapi
	29	159	20.2	2665	6	AX333017	AX333017 Sequence
	30	159	20.2	2665	9	D89077	D89077 Homo sapien
c	31	156.2	19.9	145068	9	HSDJ977B1	AL050318 Human DNA
c	32	156.2	19.9	145833	2	AC026539	AC026539 Homo sapi
	33	141.8	18.0	432	6	CQ051968	CQ051968 Sequence
	34	141.8	18.0	432	6	CQ067012	CQ067012 Sequence
	35	141.8	18.0	432	6	CQ094063	CQ094063 Sequence
	36	141.8	18.0	432	6	CQ132849	CQ132849 Sequence
	37	141.8	18.0	432	6	CQ171433	CQ171433 Sequence
	38	141.8	18.0	432	6	CQ200546	CQ200546 Sequence
	39	141.8	18.0	432	6	CQ216063	CQ216063 Sequence
	40	141.8	18.0	432	6	CQ254624	CQ254624 Sequence
	41	141.8	18.0	432	6	CQ291709	CQ291709 Sequence
	42	141.8	18.0	432	6	CQ328769	CQ328769 Sequence
	43	141.8	18.0	448	6	CQ068653	CQ068653 Sequence
	44	141.8	18.0	448	6	CQ293385	CQ293385 Sequence
	45	139.2	17.7	849	5	AY278230	AY278230 Gallus ga

ALIGNMENTS

RESULT 1
AX511153
LOCUS AX511153 786 bp DNA linear PAT 27-SEP-2002
DEFINITION Sequence 4 from Patent WO0242452.
ACCESSION AX511153
VERSION AX511153.1 GI:23392046
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Mcglade,J.C. and Loreto,M.P.
TITLE Adapter gene
JOURNAL Patent: WO 0242452-A 4 30-MAY-2002;
The Hospital for Sick Children (CA)
FEATURES Location/Qualifiers
source 1. .786
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN

Query Match 100.0%; Score 786; DB 6; Length 786;
Best Local Similarity 100.0%; Pred. No. 8.5e-195;
Matches 786; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
```

```
Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780

Qy        781 GCCTAG 786
              ||||||
Db        781 GCCTAG 786
```

RESULT 2
AX572845

```
LOCUS       AX572845                786 bp    DNA     linear   PAT 29-NOV-2002
DEFINITION  Sequence 1 from Patent WO02055707.
ACCESSION   AX572845
VERSION     AX572845.1  GI:26004935
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Holland,S.J., Mendenhall,M.K., Pardo,J., Spencer,C., Fu,A.C.,
            Luo,Y., Payan,D.G., Mancebo,H.S., Wu,J., Zhou,X., Shen,M.,
            Liao,X.C. and Sheng,N.
  TITLE     Cloning of an inhibitor of antigen-receptor signaling by a
            retroviral-based functional screen
  JOURNAL   Patent: WO 02055707-A 1 18-JUL-2002;
            Rigel Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .786
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS             1. .786
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD58542.1"
                     /db_xref="GI:26004936"
                     /translation="MGSLPSRRKSLPSPSLSSSVQGQGPVTMEAERSKATAVALGSFP
                     AGGPAELSLRLGEPLTIVSEDGDWWTVLSEVSGREYNIPSVHVAKVSHGWLYEGLSRE
                     KAEELLLLPGNPGGAFLIRESQTRRGSYSLSVRLSRPASWDRIRHYRIHCLDNGWLYI
                     SPRLTFPSLQALVDHYSELADDICCLLKEPCVLQRAGPLPGKDIPLPVTVQRTPLNWK
                     ELDSSLLFSEAATGEESLLSEGLRESLSFYISLNDEAVSLDDA"
ORIGIN
```

```
  Query Match             100.0%;  Score 786;  DB 6;  Length 786;
  Best Local Similarity   100.0%;  Pred. No. 8.5e-195;
  Matches  786;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780

Qy        781 GCCTAG 786
              ||||||
Db        781 GCCTAG 786
```

RESULT 3
AF290985

```
LOCUS       AF290985                786 bp    mRNA    linear   PRI 21-JAN-2003
DEFINITION  Homo sapiens Src-like adaptor protein-2 mRNA, complete cds.
ACCESSION   AF290985
VERSION     AF290985.1  GI:17351920
KEYWORDS    .
```

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 786)
  AUTHORS   Loreto,M.P. and McGlade,C.J.
  TITLE     Cloning and characterization of human Src-like adaptor protein 2
            and a novel splice isoform, SLAP-2-v
  JOURNAL   Oncogene 22 (2), 266-273 (2003)
  MEDLINE   22415750
   PUBMED   12527895
REFERENCE   2  (bases 1 to 786)
  AUTHORS   Loreto,M.P. and McGlade,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-2000) Brain Tumour Research Centre, Hospital for
            Sick Children, 555 University Avenue, Toronto, Ont M5G 1X8, Canada
FEATURES             Location/Qualifiers
     source          1. .786
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /tissue_type="thymus"
     CDS             1. .786
                     /note="SLAP-2"
                     /codon_start=1
                     /product="Src-like adaptor protein-2"
                     /protein_id="AAL38197.1"
                     /db_xref="GI:17351921"
                     /translation="MGSLPSRRKSLPSPSLSSSVQGQGPVTMEAERSKATAVALGSFP
                     AGGPAELSLRLGEPLTIVSEDGDWWTVLSEVSGREYNIPSVHVAKVSHGWLYEGLSRE
                     KAEELLLLPGNPGGAFLIRESQTRRGSYSLSVRLSRPASWDRIRHYRIHCLDNGWLYI
                     SPRLTFPSLQALVDHYSELADDICCLLKEPCVLQRAGPLPGKDIPLPVTVQRTPLNWK
                     ELDSSLLFSEAATGEESLLSEGLRESLSFYISLNDEAVSLDDA"
ORIGIN

  Query Match           100.0%;  Score 786;  DB 9;  Length 786;
  Best Local Similarity  100.0%;  Pred. No. 8.5e-195;
  Matches  786;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780

Qy        781 GCCTAG 786
              ||||||
Db        781 GCCTAG 786


RESULT 4
AF326353
LOCUS       AF326353                786 bp    mRNA    linear   PRI 08-NOV-2001
DEFINITION  Homo sapiens Src-like adapter protein-2 mRNA, complete cds.
ACCESSION   AF326353
VERSION     AF326353.1  GI:16797891
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 786)
  AUTHORS   Holland,S.J., Liao,X.C., Mendenhall,M.K., Zhou,X., Pardo,J.,
            Chu,P., Spencer,C., Fu,A.C., Sheng,N., Yu,P., Pali,E., Nagin,A.,
            Shen,M., Yu,S., Chan,E., Wu,X., Li,C., Woisetschlager,M.,
            Aversa,G., Kolbinger,F., Bennett,M.K., Molineaux,S., Luo,Y.,
            Payan,D.G., Mancebo,H.S.Y. and Wu,J.
  TITLE     Functional Cloning of Src-like Adapter Protein-2 (SLAP-2), a Novel
            Inhibitor of Antigen Receptor Signaling
  JOURNAL   J. Exp. Med. 194 (9), 1263-1276 (2001)
  MEDLINE   21553259
   PUBMED   11696592
REFERENCE   2  (bases 1 to 786)
  AUTHORS   Holland,S.J., Mendenhall,M.K., Zhou,X., Spencer,C., Pardo,J.,
            Fu,A.C., Sheng,N., Shen,M., Liao,C., Luo,Y., Payan,D.G.,
            Mancebo,H.S.Y. and Wu,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2000) Rigel Pharmaceutical Inc., 240 East Grand
            Avenue, South San Francisco, CA 94080, USA
FEATURES             Location/Qualifiers
     source          1. .786
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="20"
     CDS             1. .786
                     /note="SLAP2; SH2 and SH3 domain-containing adaptor"
                     /codon_start=1
                     /product="Src-like adapter protein-2"
                     /protein_id="AAL29204.1"
                     /db_xref="GI:16797892"
                     /translation="MGSLPSRRKSLPSPSLSSSVQGQGPVTMEAERSKATAVALGSFP
                     AGGPAELSLRLGEPLTIVSEDGDWWTVLSEVSGREYNIPSVHVAKVSHGWLYEGLSRE
                     KAEELLLLPGNPGGAFLIRESQTRRGSYSLSVRLSRPASWDRIRHYRIHCLDNGWLYI
                     SPRLTFPSLQALVDHYSELADDICCLLKEPCVLQRAGPLPGKDIPLPVTVQRTPLNWK
                     ELDSSLLFSEAATGEESLLSEGLRESLSFYISLNDEAVSLDDA"
ORIGIN

  Query Match           100.0%;  Score 786;  DB 9;  Length 786;
```

Best Local Similarity 100.0%; Pred. No. 8.5e-195;
Matches 786; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780

Qy        781 GCCTAG 786
              ||||||
Db        781 GCCTAG 786
```

RESULT 5
AK025645
LOCUS AK025645 2415 bp mRNA linear PRI 13-SEP-2003
DEFINITION Homo sapiens cDNA: FLJ21992 fis, clone HEP06554.
ACCESSION AK025645
VERSION AK025645.1 GI:10438227
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2415)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:flcdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Departent of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
FEATURES Location/Qualifiers
source 1. .2415
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HEP06554"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone_lib="HEP"
/note="cloning vector pME18SFL3"
CDS 61. .846
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB15201.1"
/db_xref="GI:10438228"
/translation="MGSLPSRRKSLPSPSLSSSVQGQGPVTMEAERSKATAVALGSFP
AGGPAELSLRLGEPLTIVSEDGDWWTVLSEVSGREYNIPSVHVAKVSHGWLYEGLSRE
KAEELLLLPGNPGGAFLIRESQTRRGSYSLSVRLSRPASWDRIRHYRIHCLDNGWLYI
SPRLTFPSLQALVDHYSELADDICCLLKEPCVLQRAGPLPGKDIPLPVTVQRTPLNWK
ELDSSLLFSEAATGEESLLSEGLRESLSFYISLNDEAVSLDDA"
ORIGIN

Query Match 100.0%; Score 786; DB 9; Length 2415;
Best Local Similarity 100.0%; Pred. No. 8.2e-195;
Matches 786; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 120

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 180

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 240

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 300

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 360

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 420

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 480
```

```
Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 540

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 600

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 660

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 720

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 780

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 840

Qy        781 GCCTAG 786
              ||||||
Db        841 GCCTAG 846
```

RESULT 6
BC042041

LOCUS	BC042041 2538 bp mRNA linear PRI 30-JUN-2004
DEFINITION	Homo sapiens Src-like-adaptor 2, transcript variant 1, mRNA (cDNA clone MGC:49845 IMAGE:4429896), complete cds.
ACCESSION	BC042041
VERSION	BC042041.1 GI:27469842
KEYWORDS	MGC.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 2538)
AUTHORS	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED	12477932
REFERENCE	2 (bases 1 to 2538)
AUTHORS	Strausberg,R.
TITLE	Direct Submission
JOURNAL	Submitted (23-DEC-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov

```
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 88 Row: a Column: 20
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 28416422.
FEATURES              Location/Qualifiers
     source           1. .2538
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="MGC:49845 IMAGE:4429896"
                      /tissue_type="Prostate, adenocarcinoma."
                      /clone_lib="NIH_MGC_91"
                      /lab_host="DH10B"
                      /note="Vector: pCMV-SPORT6"
     gene             1. .2538
                      /gene="SLA2"
                      /note="synonyms: SLAP-2, MGC49845, FLJ21992, SLAP2,
                      C20orf156"
                      /db_xref="LocusID:84174"
                      /db_xref="MIM:606577"
     CDS              363. .1148
                      /gene="SLA2"
                      /codon_start=1
                      /product="Src-like-adaptor 2, isoform a"
                      /protein_id="AAH42041.1"
                      /db_xref="GI:27469843"
                      /db_xref="LocusID:84174"
                      /db_xref="MIM:606577"
                      /translation="MGSLPSRRKSLPSPSLSSSVQGQGPVTMEAERSKATAVALGSFP
                      AGGPAELSLRLGEPLTIVSEDGDWWTVLSEVSGREYNIPSVHVAKVSHGWLYEGLSRE
                      KAEELLLLPGNPGGAFLIRESQTRRGSYSLSVRLSRPASWDRIRHYRIHCLDNGWLYI
                      SPRLTFPSLQALVDHYSELADDICCLLKEPCVLQRAGPLPGKDIPLPVTVQRTPLNWK
                      ELDSSLLFSEAATGEESLLSEGLRESLSFYISLNDEAVSLDDA"
ORIGIN
```

Query Match 100.0%; Score 786; DB 9; Length 2538;
Best Local Similarity 100.0%; Pred. No. 8.2e-195;
Matches 786; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 422

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 482

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 542

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 602

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 662

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
```

```
Db        663 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 722

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
Db        723 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 782

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
Db        783 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 842

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
Db        843 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 902

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
Db        903 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 962

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
Db        963 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 1022

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
Db       1023 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 1082

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780
Db       1083 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 1142

Qy        781 GCCTAG 786
Db       1143 GCCTAG 1148
```

```
RESULT 7
AX452880
LOCUS       AX452880               2567 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 1 from Patent WO0242457.
ACCESSION   AX452880
VERSION     AX452880.1  GI:21712520
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Chang,H., Yang,W.P., Wu,Y., Whitney,G.S., Perez-Villar,J.J. and
            Kanner,S.B.
  TITLE     Cloning and expression of human slap-2: a novel sh2/sh3
            domain-containing human slap homologue having immune cell-specific
            expression
  JOURNAL   Patent: WO 0242457-A 1 30-MAY-2002;
            Bristol-Myers Squibb Co. (US)
FEATURES             Location/Qualifiers
     source          1. .2567
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             100.0%;  Score 786;  DB 6;  Length 2567;
  Best Local Similarity   100.0%;  Pred. No. 8.2e-195;
  Matches  786;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
Db        415 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 474

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
Db        475 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 534
```

```
Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
Db        535 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 594

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
Db        595 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 654

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
Db        655 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 714

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
Db        715 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 774

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
Db        775 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 834

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
Db        835 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 894

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
Db        895 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 954

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
Db        955 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 1014

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
Db       1015 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 1074

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
Db       1075 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 1134

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780
Db       1135 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 1194

Qy        781 GCCTAG 786
Db       1195 GCCTAG 1200
```

```
RESULT 8
AX780857
LOCUS       AX780857               2788 bp    DNA     linear   PAT 14-JUL-2003
DEFINITION  Sequence 3014 from Patent WO03039443.
ACCESSION   AX780857
VERSION     AX780857.1  GI:32697851
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Haferlach,T., Schoch,C., Kern,W., Kohlmann,A., Schnittger,S.,
            Dugas,M., Eils,R., Brors,B. and Mergenthaler,S.
  TITLE     Novel genetic markers for leukemias
  JOURNAL   Patent: WO 03039443-A 3014 15-MAY-2003;
            Deutsches Krebsforschungszentrum (DE) ;
            Ludwig-Maximilian-Universitaet Muenchen (DE) ; Haferlach, Torsten,
            PD Dr. Dr. (DE) ; Schoch, Claudia (DE) ; Kern, Wolfgang (DE)
FEATURES             Location/Qualifiers
     source          1. .2788
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
```

```
                     /db_xref="taxon:9606"
ORIGIN

  Query Match           99.9%;  Score 785;  DB 6;  Length 2788;
  Best Local Similarity   99.9%;  Pred. No. 1.5e-194;
  Matches  785;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 446

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 CAAGGCCANGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 506

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 566

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        567 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 626

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        627 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 686

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        687 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 746

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        747 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 806

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        807 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 866

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        867 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 926

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        927 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 986

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        987 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 1046

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1047 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 1106

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1107 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 1166

Qy        781 GCCTAG 786
              ||||||
Db       1167 GCCTAG 1172


RESULT 9
AX443133
LOCUS       AX443133               1183 bp    DNA     linear   PAT 02-JUL-2002
DEFINITION  Sequence 74 from Patent WO0216599.
ACCESSION   AX443133
VERSION     AX443133.1  GI:21690555
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Burgess,C.E., Conley,P.B., Grosse,W.M., Hart,M., Kekuda,R.,
            Shimkets,R.A., Spytek,K.A., Szekeres,E.S., Tomlinson,J.E.,
            Topper,J.N. and Yang,R.B.
  TITLE     Proteins and nucleic acids encoding same
  JOURNAL   Patent: WO 0216599-A 74 28-FEB-2002;
            Curagen Corporation (US) ; COR THERAPEUTICS, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .1183
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match           99.8%;  Score 784.4;  DB 6;  Length 1183;
  Best Local Similarity   99.9%;  Pred. No. 2.2e-194;
  Matches  785;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        398 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 457

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        458 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 517

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        518 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 577

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        578 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 637

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        638 AACATCCCCAGCGTCCACGTGGGCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 697

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        698 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 757

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        758 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 817

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        818 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 877

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        878 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 937

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        938 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 997

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        998 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 1057

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1058 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 1117

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1118 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 1177

Qy        781 GCCTAG 786
```

```
                ||||||
Db       1178 GCCTAG 1183


RESULT 10
AX443135/c
LOCUS       AX443135                1183 bp    DNA     linear   PAT 02-JUL-2002
DEFINITION  Sequence 76 from Patent WO0216599.
ACCESSION   AX443135
VERSION     AX443135.1  GI:21690556
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Burgess,C.E., Conley,P.B., Grosse,W.M., Hart,M., Kekuda,R.,
            Shimkets,R.A., Spytek,K.A., Szekeres,E.S., Tomlinson,J.E.,
            Topper,J.N. and Yang,R.B.
  TITLE     Proteins and nucleic acids encoding same
  JOURNAL   Patent: WO 0216599-A 76 28-FEB-2002;
            Curagen Corporation (US) ; COR THERAPEUTICS, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .1183
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match           99.8%;  Score 784.4;  DB 6;  Length 1183;
  Best Local Similarity 99.9%;  Pred. No. 2.2e-194;
  Matches  785;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        786 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 727

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 667

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 607

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 547

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        546 AACATCCCCAGCGTCCACGTGGGCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 487

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 427

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 367

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 307

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 247

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 187
```

```
Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 127

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 67

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 7

Qy        781 GCCTAG 786
              ||||||
Db          6 GCCTAG 1


RESULT 11
AX511155
LOCUS       AX511155                  737 bp    DNA     linear   PAT 27-SEP-2002
DEFINITION  Sequence 6 from Patent WO0242452.
ACCESSION   AX511155
VERSION     AX511155.1  GI:23392047
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Mcglade,J.C. and Loreto,M.P.
  TITLE     Adapter gene
  JOURNAL   Patent: WO 0242452-A 6 30-MAY-2002;
            The Hospital for Sick Children (CA)
FEATURES             Location/Qualifiers
     source          1. .737
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match           84.4%;  Score 663.4;  DB 6;  Length 737;
  Best Local Similarity 93.4%;  Pred. No. 1.1e-162;
  Matches  735;  Conservative    0;  Mismatches    1;  Indels   51;  Gaps    2;

Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
```

```
Db        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCT--------- 531

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
                                                  ||||||||||||||||||||
Db        532 ----------------------------------------GAGGGCTGGCCCGCTCCCT 550

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        551 GGCAAGGATATACCCCTACCTGTGACTGTGCGGAGGACACCACTCAACTGGAAAGAGCTG 610

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        611 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 670

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGA-CGAGGCTGTCTCTTTGGATGA 779
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        671 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGAGCGAGGCTGTCTCTTTGGATGA 730

Qy        780 TGCCTAG 786
              |||||||
Db        731 TGCCTAG 737
```

RESULT 12
AF290986
LOCUS AF290986 737 bp mRNA linear PRI 21-JAN-2003
DEFINITION Homo sapiens Src-like adaptor protein-2 splice isoform mRNA, complete cds; alternatively spliced.
ACCESSION AF290986
VERSION AF290986.1 GI:17351922
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 737)
AUTHORS Loreto,M.P. and McGlade,C.J.
TITLE Cloning and characterization of human Src-like adaptor protein 2 and a novel splice isoform, SLAP-2-v
JOURNAL Oncogene 22 (2), 266-273 (2003)
MEDLINE 22415750
PUBMED 12527895
REFERENCE 2 (bases 1 to 737)
AUTHORS Loreto,M.P. and McGlade,C.J.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-2000) Brain Tumour Research Centre, Hospital for Sick Children, 555 University Avenue, Toronto, Ont M5G 1X8, Canada
FEATURES Location/Qualifiers
source 1. .737
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="20"
/tissue_type="thymus"
CDS 1. .633
/note="SLAP-2-v"
/codon_start=1
/product="Src-like adaptor protein-2 splice isoform"
/protein_id="AAL38198.1"
/db_xref="GI:17351923"
/translation="MGSLPSRRKSLPSPSLSSSVQGQGPVTMEAERSKATAVALGSFP
AGGPAELSLRLGEPLTIVSEDGDWWTVLSEVSGREYNIPSVHVAKVSHGWLYEGLSRE
KAEELLLLPGNPGGAFLIRESQTRRGSYSLSVRLSRPASWDRIRHYRIHCLDNGWLYI
SPRLTFPSLQALVDHYSEGWPAPWQGYTPTCDCAEDTTQLERAGQLPPVF"
ORIGIN

Query Match 84.4%; Score 663.4; DB 9; Length 737;
Best Local Similarity 93.4%; Pred. No. 1.1e-162;

Matches 735; Conservative 0; Mismatches 1; Indels 51; Gaps 2;

```
Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCT--------- 531

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
                                                  ||||||||||||||||||||
Db        532 ----------------------------------------GAGGGCTGGCCCGCTCCCT 550

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        551 GGCAAGGATATACCCCTACCTGTGACTGTGCGGAGGACACCACTCAACTGGAAAGAGCTG 610

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        611 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 670

Qy        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGA-CGAGGCTGTCTCTTTGGATGA 779
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        671 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGAGCGAGGCTGTCTCTTTGGATGA 730

Qy        780 TGCCTAG 786
              |||||||
Db        731 TGCCTAG 737
```

RESULT 13
BC052655
LOCUS BC052655 1384 bp mRNA linear ROD 04-NOV-2003
DEFINITION Mus musculus Src-like-adaptor 2, mRNA (cDNA clone MGC:60811 IMAGE:30040401), complete cds.
ACCESSION BC052655
VERSION BC052655.1 GI:30851667
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1384)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,

```
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1384)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Drs. Dennis Taub, Dan Longo (NIA, USA),
            Jonathan Keller (NCI, USA)
            cDNA Library Preparation: Yulan Piao and Minoru Ko (National
            Institute on Aging, NIH: http://lgsun.grc.nia.nih.gov/cDNA/)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 112 Row: e Column: 17
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES            Location/Qualifiers
     source         1. .1384
                    /organism="Mus musculus"
                    /mol_type="mRNA"
                    /strain="C57BL/6NCr"
                    /db_xref="taxon:10090"
                    /clone="MGC:60811 IMAGE:30040401"
                    /tissue_type="Hematopoietic Stem Cell,
                    (Lin-/c-Kit+/Sca-1-), mouse, 10 wks"
                    /clone_lib="NIA Mouse Hematopoietic Stem Cell
                    (Lin-/c-Kit+/Sca-1-) cDNA Library (Long)"
                    /lab_host="DH10B"
                    /note="Vector: pSPORT1"
     gene           1. .1384
                    /gene="Sla2"
                    /note="synonyms: SLAP2, SLAP-2"
                    /db_xref="LocusID:77799"
                    /db_xref="MGI:1925049"
     CDS            330. .1109
                    /gene="Sla2"
                    /codon_start=1
                    /product="Sla2 protein"
                    /protein_id="AAH52655.1"
                    /db_xref="GI:30851668"
                    /db_xref="LocusID:77799"
                    /db_xref="MGI:1925049"
                    /translation="MGSLSSRGKTSSPSPSSSGPDQEPVSMQPERRKVTAVALGSFPA
                    GEQARLSLRLGEPLTIISEDGDWWTVQSEVSGREYHMPSVYVAKVAHGWLYEGLSREK
                    AEELLLLPGNPGGAFLIRESQTRRGCYSLSVRLSRPASWDRIRHYRIQRLDNGWLYIS
                    PRLTFPSLHALVEHYSELADGICCPLREPCVLQKLGPLPGKDTPPPVTVPTSSLNWKK
                    LDRSLLFLEAPASGEASLLSEGLRESLSSYISLAEDPLDDA"
     misc_feature   432. .596
                    /gene="Sla2"
                    /note="SH3; Region: SH3 domain. SH3 (Src homology 3)
                    domains are often indicative of a protein involved in
                    signal transduction related to cytoskeletal organisation.
                    First described in the Src cytoplasmic tyrosine kinase.
                    The structure is a partly opened beta barrel"
                    /db_xref="CDD:pfam00018"
     misc_feature   603. .899
                    /gene="Sla2"
                    /note="SH2; Region: Src homology 2 domains"
                    /db_xref="CDD:cd00173"
ORIGIN

  Query Match             68.6%;  Score 539.4;  DB 10;  Length 1384;
  Best Local Similarity   82.0%;  Pred. No. 3.1e-130;
  Matches  647;  Conservative    0;  Mismatches  136;  Indels    6;  Gaps    2;

Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              |||||||||  || |||||||| | ||| |     | |||||  |||   || ||||||| 
Db        330 ATGGGAAGTTTGTCCAGCAGAGGGAAAAC---CTCCAGCCCCAGCCCCAGCTCCTCTGGT 386

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              | ||  |||||  |||  ||| |||||  ||  ||||  |||  ||||||  ||||||  |||||||||
Db        387 CCAGACCAGGAACCCGTGTCCATGCAACCAGAAAGACGCAAGGTCACAGCTGTGGCCCTG 446

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||  |||||||   |  ||||    ||  ||  ||||||||||||||||||  |||||
Db        447 GGCAGTTTCCCAGCAGGTGAACAGGCCAGACTATCTCTGAGACTCGGGGAGCCGCTGACC 506

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              |||  |||||||||||||||||  |||||||||  ||  |  |||  ||||||||||||||||||||
Db        507 ATCATCTCTGAGGATGGAGATTGGTGGACAGTCCAGTCGGAAGTCTCAGGCAGAGAGTAC 566

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||  |||||  ||  |  |||||  ||||||  ||||  ||||||||||  ||||||||||||
Db        567 CACATGCCCAGTGTGTATGTGGCTAAAGTCGCCCACGGGTGGCTGTACGAGGGCCTGAGC 626

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||  ||||||||  ||  ||||||||||||||||  ||||||||||||||||||
Db        627 CGGGAGAAAGCCGAGGAACTACTCCTGTTACCTGGGAACCCCGGAGGGGCCTTCCTCATC 686

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||  ||  ||  ||||  ||||  ||||||||||||||||
Db        687 CGGGAGAGCCAGACCAGGAGAGGCTGCTATTCCCTGTCCGTCCGACTCAGCCGCCCTGCA 746

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||  |||||||||||||||||||||||||||||||  ||  |  ||||||||||||||||||||||||
Db        747 TCTTGGGACCGGATCAGACACTACAGGATACAGCGTCTTGACAATGGCTGGCTGTACATC 806

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              |||||  |||||||||||||||||||||||  |||  |||||||  ||||||||||||  ||
Db        807 TCACCTCGCCTCACCTTCCCCTCACTCCACGCCTTGGTGGAGCATTACTCTGAGCTAGCA 866

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              ||||  |||||||||  |  ||||  ||||||  ||||||||||||  |  |||  ||  ||  |||
Db        867 GATGGCATCTGCTGTCCCCTCAGGGAGCCGTGTGTCCTGCAGAAGCTTGGGCCACTACCT 926
```

```
Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||| |||| ||| | ||||||||||||||   |    || |||| || |||||| |||||
Db        927 GGCAAAGATACACCTCCACCTGTGACTGTGCCAACATCATCACTAAATTGGAAAAAGCTG 986

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAG---CTGCCACAGGGGAGGAGTCTCTTCTCAGTGAG 717
              ||| ||  ||||||||||   ||||   |||| |  |||||||  ||||| |||||||||
Db        987 GACCGCAGCCTCCTGTTTCTGGAAGCACCTGCGAGTGGGGAGGCATCTCTGCTCAGTGAG 1046

Qy        718 GGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGAT 777
              || ||||| |||||||||||| | |||||||||||||  ||| ||  |  |   |  | |
Db       1047 GGGCTCCGAGAGTCCCTCAGTTCCTACATCAGCCTGGCTGAGGACCCCTTGGATGATGCT 1106

Qy        778 GATGCCTAG 786
                |  ||| |
Db       1107 TAGCCCTGG 1115
```

RESULT 14
AX511150
LOCUS AX511150 1348 bp DNA linear PAT 27-SEP-2002
DEFINITION Sequence 1 from Patent WO0242452.
ACCESSION AX511150
VERSION AX511150.1 GI:23392044
KEYWORDS .
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Mcglade,J.C. and Loreto,M.P.
TITLE Adapter gene
JOURNAL Patent: WO 0242452-A 1 30-MAY-2002;
The Hospital for Sick Children (CA)
FEATURES Location/Qualifiers
source 1. .1348
/organism="Mus musculus"
/mol_type="unassigned DNA"
/db_xref="taxon:10090"
ORIGIN

Query Match 68.4%; Score 537.8; DB 6; Length 1348;
Best Local Similarity 81.9%; Pred. No. 8e-130;
Matches 646; Conservative 0; Mismatches 137; Indels 6; Gaps 2;

```
Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||| || |||||||| | ||| |     | ||||| |||   || |||||||
Db        282 ATGGGAAGTTTGTCCAGCAGAGGGAAAAC---CTCCAGCCCCAGCCCCAGCTCCTCTGGT 338

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              | || ||||| ||| ||| ||||| || |||| |||  |||||  |||||| |||||||||
Db        339 CCAGACCAGGAACCCGTGTCCATGCAACCAGAAAGACACAAGGTCACAGCTGTGGCCCTG 398

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              |||||||||||| |||||||  | ||||   || || ||||||||||||||||  |||||
Db        399 GGCAGTTTCCCAGCAGGTGAACAGGCCAGACTATCTCTGAGACTCGGGGAGCCGCTGACC 458

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||| |||||||||||||||| |||||||| || | ||| ||||||||||||||||||||
Db        459 ATCATCTCTGAGGATGGAGATTGGTGGACAGTCCAGTCGGAAGTCTCAGGCAGAGAGTAC 518

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              |||| ||||| ||  | ||||| |||||| |||| |||||||||| ||||||||||||
Db        519 CACATGCCCAGTGTGTATGTGGCTAAAGTCGCCCACGGGTGGCTGTACGAGGGCCTGAGC 578

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              |||||||||| |||||||| ||  |||||||||||||||| ||||||||||||||||||
Db        579 CGGGAGAAAGCCGAGGAACTACTCCTGTTACCTGGGAACCCCGGAGGGGCCTTCCTCATC 638

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||  || || ||||| ||||| ||||||||||||||||
Db        639 CGGGAGAGCCAGACCAGGAGAGGCTGCTATTCCCTGTCCGTCCGACTCAGCCGCCCTGCA 698

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              || ||||||||||||||||||||||||||| ||  | |||||||||||||||||||||||
Db        699 TCTTGGGACCGGATCAGACACTACAGGATACAGCGTCTTGACAATGGCTGGCTGTACATC 758

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||| |||||||||||||||||||||||| ||| |||||||| |||||||||||||| ||
Db        759 TCACCTCGCCTCACCTTCCCCTCACTCCACGCCTTGGTGGAGCATTACTCTGAGCTAGCA 818

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              |||| |||||||||  |  |||| |||||| |||||||||||||  | ||| || || |||
Db        819 GATGGCATCTGCTGTCCCCTCAGGGAGCCGTGTGTCCTGCAGAAGCTTGGGCCACTACCT 878

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||| |||| ||| | |||||||||||||||  |    || |||| || |||||| |||||
Db        879 GGCAAAGATACACCTCCACCTGTGACTGTGCCAACATCATCACTAAATTGGAAAAAGCTG 938

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAG---CTGCCACAGGGGAGGAGTCTCTTCTCAGTGAG 717
              ||| ||  ||||||||||   ||||   |||| |  |||||||  ||||| |||||||||
Db        939 GACCGCAGCCTCCTGTTTCTGGAAGCACCTGCGAGTGGGGAGGCATCTCTGCTCAGTGAG 998

Qy        718 GGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGAT 777
              || ||||| |||||||||||| | |||||||||||||  ||| ||  |  |   |  | |
Db        999 GGGCTCCGAGAGTCCCTCAGTTCCTACATCAGCCTGGCTGAGGACCCCTTGGATGATGCT 1058

Qy        778 GATGCCTAG 786
                |  ||| |
Db       1059 TAGCCCTGG 1067
```

RESULT 15
AF287467
LOCUS AF287467 1348 bp mRNA linear ROD 03-JUN-2002
DEFINITION Mus musculus Src-like adaptor protein-2 mRNA, complete cds.
ACCESSION AF287467
VERSION AF287467.1 GI:17351918
KEYWORDS .
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1348)
AUTHORS Loreto,M.P., Berry,D.M. and McGlade,C.J.
TITLE Functional cooperation between c-Cbl and Src-like adaptor protein 2
in the negative regulation of T-cell receptor signaling
JOURNAL Mol. Cell. Biol. 22 (12), 4241-4255 (2002)
MEDLINE 22022020
PUBMED 12024036
REFERENCE 2 (bases 1 to 1348)
AUTHORS Loreto,M.P. and McGlade,C.J.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-2000) Brain Tumour Research Centre, Hospital for
Sick Children, 555 University Avenue, Toronto, Ont M5G 1X8, Canada
FEATURES Location/Qualifiers
source 1. .1348
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/dev_stage="day 15 embryo"
CDS 282. .1061
/note="SLAP-2"
/codon_start=1
/product="Src-like adaptor protein-2"
/protein_id="AAL38196.1"
/db_xref="GI:17351919"
/translation="MGSLSSRGKTSSPSPSSSGPDQEPVSMQPERHKVTAVALGSFPA
GEQARLSLRLGEPLTIISEDGDWWTVQSEVSGREYHMPSVYVAKVAHGWLYEGLSREK
AEELLLLPGNPGGAFLIRESQTRRGCYSLSVRLSRPASWDRIRHYRIQRLDNGWLYIS
PRLTFPSLHALVEHYSELADGICCPLREPCVLQKLGPLPGKDTPPPVTVPTSSLNWKK
LDRSLLFLEAPASGEASLLSEGLRESLSSYISLAEDPLDDA"
misc_feature 381. .551
/note="Region: SH3 domain"
misc_feature 552. .824
/note="Region: SH2 domain"
polyA_signal 1301. .1306

ORIGIN

Query Match 68.4%; Score 537.8; DB 10; Length 1348;
Best Local Similarity 81.9%; Pred. No. 8e-130;
Matches 646; Conservative 0; Mismatches 137; Indels 6; Gaps 2;

```
Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              |||||||||  ||  |||||||||  |  |||  |     |  |||||  |||    ||  |||||||
Db        282 ATGGGAAGTTTGTCCAGCAGAGGGAAAAC---CTCCAGCCCCAGCCCCAGCTCCTCTGGT 338

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              |  ||  |||||  |||  |||  |||||  ||  ||||  |||   |||||  ||||||  |||||||||
Db        339 CCAGACCAGGAACCCGTGTCCATGCAACCAGAAAGACACAAGGTCACAGCTGTGGCCCTG 398

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||  |||||||   |  ||||    ||  ||  ||||||||||||||||  |||||
Db        399 GGCAGTTTCCCAGCAGGTGAACAGGCCAGACTATCTCTGAGACTCGGGGAGCCGCTGACC 458

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              |||  ||||||||||||||||  ||||||||  ||  |  |||  ||||||||||||||||||||
Db        459 ATCATCTCTGAGGATGGAGATTGGTGGACAGTCCAGTCGGAAGTCTCAGGCAGAGAGTAC 518

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||  |||||  ||   |  |||||  ||||||  ||||  ||||||||||  ||||||||||||
Db        519 CACATGCCCAGTGTGTATGTGGCTAAAGTCGCCCACGGGTGGCTGTACGAGGGCCTGAGC 578

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||  ||||||||  ||   ||||||||||||||||  ||||||||||||||||||||
Db        579 CGGGAGAAAGCCGAGGAACTACTCCTGTTACCTGGGAACCCCGGAGGGGCCTTCCTCATC 638

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||  ||  ||  |||||  |||||  ||||||||||||||||
Db        639 CGGGAGAGCCAGACCAGGAGAGGCTGCTATTCCCTGTCCGTCCGACTCAGCCGCCCTGCA 698

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||  ||||||||||||||||||||||||||  ||   |  ||||||||||||||||||||||||||||
Db        699 TCTTGGGACCGGATCAGACACTACAGGATACAGCGTCTTGACAATGGCTGGCTGTACATC 758

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              |||||  ||||||||||||||||||||||  |||  |||||||  ||||||||||||||  ||
Db        759 TCACCTCGCCTCACCTTCCCCTCACTCCACGCCTTGGTGGAGCATTACTCTGAGCTAGCA 818

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              ||||  |||||||||  |  ||||  ||||||  ||||||||||||  |  |||  ||  ||  |||
Db        819 GATGGCATCTGCTGTCCCCTCAGGGAGCCGTGTGTCCTGCAGAAGCTTGGGCCACTACCT 878

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              |||||  ||||  |||  |  ||||||||||||||  |    ||  ||||  ||  ||||||  |||||
Db        879 GGCAAAGATACACCTCCACCTGTGACTGTGCCAACATCATCACTAAATTGGAAAAAGCTG 938

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAG---CTGCCACAGGGGAGGAGTCTCTTCTCAGTGAG 717
              |||  ||  ||||||||||    ||||    ||||  |  |||||||   |||||  ||||||||||
Db        939 GACCGCAGCCTCCTGTTTCTGGAAGCACCTGCGAGTGGGGAGGCATCTCTGCTCAGTGAG 998

Qy        718 GGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGAT 777
              ||  |||||  ||||||||||||  |  ||||||||||||  |||  ||  |  |  |  |  | |
Db        999 GGGCTCCGAGAGTCCCTCAGTTCCTACATCAGCCTGGCTGAGGACCCCTTGGATGATGCT 1058

Qy        778 GATGCCTAG 786
                |  |||  |
Db       1059 TAGCCCTGG 1067
```

Search completed: November 16, 2004, 22:36:22
Job time : 3774 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 16, 2004, 23:32:28 ; Search time 4546 Seconds
(without alignments)
2715.049 Million cell updates/sec

Title: US-10-043-649-2
Perfect score: 1351
Sequence: 1 MGSLPSRRKSLPSPSLSSSV.....RESLSFYISLNDEAVSLDDA 261

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10043649/runat_16112004_060536_28719/app_query.fasta_1.455
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10043649_@CGN_1_1_3731_@runat_16112004_060536_28719 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1351	100.0	786	6	AX511153	AX511153 Sequence
2	1351	100.0	786	6	AX572845	AX572845 Sequence
3	1351	100.0	786	9	AF290985	AF290985 Homo sapi
4	1351	100.0	786	9	AF326353	AF326353 Homo sapi
5	1351	100.0	2415	9	AK025645	AK025645 Homo sapi
6	1351	100.0	2538	9	BC042041	BC042041 Homo sapi
7	1351	100.0	2567	6	AX452880	AX452880 Sequence
8	1347	99.7	1183	6	AX443133	AX443133 Sequence
c 9	1347	99.7	1183	6	AX443135	AX443135 Sequence
10	1345	99.6	2788	6	AX780857	AX780857 Sequence
11	1200.5	88.9	737	6	AX511155	AX511155 Sequence
12	1200.5	88.9	737	9	AF290986	AF290986 Homo sapi
13	1032	76.4	777	6	AX511151	AX511151 Sequence
14	1032	76.4	1348	6	AX511150	AX511150 Sequence
15	1032	76.4	1348	10	AF287467	AF287467 Mus muscu
16	1032	76.4	1384	10	BC052655	BC052655 Mus muscu
17	1023	75.7	1321	10	AF434990	AF434990 Mus muscu
18	770	57.0	441	6	CQ724864	CQ724864 Sequence
19	647	47.9	864	6	CQ414231	CQ414231 Sequence
20	587	43.4	1571	5	BC056035	BC056035 Xenopus l
21	505	37.4	1631	10	AY217759	AY217759 Rattus no
22	496	36.7	921	10	MMU29056	U29056 Mus musculu
23	496	36.7	2648	10	BC032922	BC032922 Mus muscu
24	496	36.7	2773	10	MMU131777	AJ131777 Mus muscu
25	495.5	36.7	841	10	AY079449	AY079449 Mus muscu
26	495.5	36.7	841	10	AY079450	AY079450 Mus muscu
27	491	36.3	1076	9	HSU30473	U30473 Homo sapien
28	491	36.3	1870	9	BC007042	BC007042 Homo sapi
29	491	36.3	2021	9	HSU44403	U44403 Human Src-l
30	491	36.3	2109	6	AX428893	AX428893 Sequence
31	491	36.3	2665	6	AX333017	AX333017 Sequence
32	491	36.3	2665	9	D89077	D89077 Homo sapien
33	491	36.3	2670	6	CQ725810	CQ725810 Sequence
34	491	36.3	3090	9	HSM808530	BX648382 Homo sapi
35	485.5	35.9	831	9	CR536537	CR536537 Homo sapi
c 36	482	35.7	145068	9	HSDJ977B1	AL050318 Human DNA
c 37	482	35.7	145833	2	AC026539	AC026539 Homo sapi
38	471.5	34.9	849	5	AY278230	AY278230 Gallus ga
39	384.5	28.5	2703	10	MUSLYNB	M57697 Mouse lyn B
40	379.5	28.1	1538	10	RATLYNBTYR	L14823 Rattus norv
41	379.5	28.1	2302	10	RNAF000301	AF000301 Rattus no
42	379.5	28.1	2824	10	RNAF000302	AF000302 Rattus no
43	374.5	27.7	1361	6	CQ729540	CQ729540 Sequence
44	374.5	27.7	1539	12	AY335782	AY335782 Synthetic
45	374.5	27.7	1613	9	BC075001	BC075001 Homo sapi

ALIGNMENTS

RESULT 1
AX511153
LOCUS AX511153 786 bp DNA linear PAT 27-SEP-2002
DEFINITION Sequence 4 from Patent WO0242452.
ACCESSION AX511153
VERSION AX511153.1 GI:23392046
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Mcglade,J.C. and Loreto,M.P.
TITLE Adapter gene
JOURNAL Patent: WO 0242452-A 4 30-MAY-2002;
The Hospital for Sick Children (CA)
FEATURES Location/Qualifiers
source 1. .786
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.: 1.37e-101 Length: 786
Score: 1351.00 Matches: 261
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

```
Query Match:            100.00%     Indels:     0
DB:                     6           Gaps:       0

US-10-043-649-2 (1-261) x AX511153 (1-786)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660

Qy        221 AspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGly 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720

Qy        241 LeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAspAsp 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780

Qy        261 Ala 261
              |||
Db        781 GCC 783

RESULT 2
AX572845
LOCUS       AX572845                786 bp    DNA     linear   PAT 29-NOV-2002
DEFINITION  Sequence 1 from Patent WO02055707.
ACCESSION   AX572845
VERSION     AX572845.1  GI:26004935
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Holland,S.J., Mendenhall,M.K., Pardo,J., Spencer,C., Fu,A.C.,
            Luo,Y., Payan,D.G., Mancebo,H.S., Wu,J., Zhou,X., Shen,M.,
            Liao,X.C. and Sheng,N.
  TITLE     Cloning of an inhibitor of antigen-receptor signaling by a
            retroviral-based functional screen
  JOURNAL   Patent: WO 02055707-A 1 18-JUL-2002;
            Rigel Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .786
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS             1. .786
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD58542.1"
                     /db_xref="GI:26004936"
                     /translation="MGSLPSRRKSLPSPSLSSSVQGQGPVTMEAERSKATAVALGSFP
                     AGGPAELSLRLGEPLTIVSEDGDWWTVLSEVSGREYNIPSVHVAKVSHGWLYEGLSRE
                     KAEELLLLPGNPGGAFLIRESQTRRGSYSLSVRLSRPASWDRIRHYRIHCLDNGWLYI
                     SPRLTFPSLQALVDHYSELADDICCLLKEPCVLQRAGPLPGKDIPLPVTVQRTPLNWK
                     ELDSSLLFSEAATGEESLLSEGLRESLSFYISLNDEAVSLDDA"
ORIGIN

Alignment Scores:
Pred. No.:              1.37e-101     Length:         786
Score:                  1351.00       Matches:        261
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     6             Gaps:           0

US-10-043-649-2 (1-261) x AX572845 (1-786)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
```

```
Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660

Qy        221 AspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGly 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720

Qy        241 LeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAspAsp 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780

Qy        261 Ala 261
              |||
Db        781 GCC 783
```

RESULT 3
AF290985

```
LOCUS       AF290985                 786 bp    mRNA    linear   PRI 21-JAN-2003
DEFINITION  Homo sapiens Src-like adaptor protein-2 mRNA, complete cds.
ACCESSION   AF290985
VERSION     AF290985.1  GI:17351920
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 786)
  AUTHORS   Loreto,M.P. and McGlade,C.J.
  TITLE     Cloning and characterization of human Src-like adaptor protein 2
            and a novel splice isoform, SLAP-2-v
  JOURNAL   Oncogene 22 (2), 266-273 (2003)
  MEDLINE   22415750
   PUBMED   12527895
REFERENCE   2  (bases 1 to 786)
  AUTHORS   Loreto,M.P. and McGlade,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-2000) Brain Tumour Research Centre, Hospital for
            Sick Children, 555 University Avenue, Toronto, Ont M5G 1X8, Canada
FEATURES             Location/Qualifiers
     source          1. .786
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /tissue_type="thymus"
     CDS             1. .786
                     /note="SLAP-2"
                     /codon_start=1
                     /product="Src-like adaptor protein-2"
                     /protein_id="AAL38197.1"
                     /db_xref="GI:17351921"
                     /translation="MGSLPSRRKSLPSPSLSSSVQGQGPVTMEAERSKATAVALGSFP
                     AGGPAELSLRLGEPLTIVSEDGDWWTVLSEVSGREYNIPSVHVAKVSHGWLYEGLSRE
                     KAEELLLLPGNPGGAFLIRESQTRRGSYSLSVRLSRPASWDRIRHYRIHCLDNGWLYI
                     SPRLTFPSLQALVDHYSELADDICCLLKEPCVLQRAGPLPGKDIPLPVTVQRTPLNWK
                     ELDSSLLFSEAATGEESLLSEGLRESLSFYISLNDEAVSLDDA"
ORIGIN
```

Alignment Scores:

Pred. No.:	1.37e-101	Length:	786
Score:	1351.00	Matches:	261
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-10-043-649-2 (1-261) x AF290985 (1-786)

```
Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660

Qy        221 AspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGly 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720

Qy        241 LeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAspAsp 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780

Qy        261 Ala 261
              |||
Db        781 GCC 783
```

RESULT 4
AF326353

```
LOCUS       AF326353                 786 bp    mRNA    linear   PRI 08-NOV-2001
DEFINITION  Homo sapiens Src-like adapter protein-2 mRNA, complete cds.
ACCESSION   AF326353
VERSION     AF326353.1  GI:16797891
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 786)
  AUTHORS   Holland,S.J., Liao,X.C., Mendenhall,M.K., Zhou,X., Pardo,J.,
            Chu,P., Spencer,C., Fu,A.C., Sheng,N., Yu,P., Pali,E., Nagin,A.,
            Shen,M., Yu,S., Chan,E., Wu,X., Li,C., Woisetschlager,M.,
            Aversa,G., Kolbinger,F., Bennett,M.K., Molineaux,S., Luo,Y.,
            Payan,D.G., Mancebo,H.S.Y. and Wu,J.
  TITLE     Functional Cloning of Src-like Adapter Protein-2 (SLAP-2), a Novel
            Inhibitor of Antigen Receptor Signaling
  JOURNAL   J. Exp. Med. 194 (9), 1263-1276 (2001)
```

```
  MEDLINE   21553259
   PUBMED   11696592
REFERENCE   2  (bases 1 to 786)
  AUTHORS   Holland,S.J., Mendenhall,M.K., Zhou,X., Spencer,C., Pardo,J.,
            Fu,A.C., Sheng,N., Shen,M., Liao,C., Luo,Y., Payan,D.G.,
            Mancebo,H.S.Y. and Wu,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2000) Rigel Pharmaceutical Inc., 240 East Grand
            Avenue, South San Francisco, CA 94080, USA
FEATURES             Location/Qualifiers
     source          1. .786
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="20"
     CDS             1. .786
                     /note="SLAP2; SH2 and SH3 domain-containing adaptor"
                     /codon_start=1
                     /product="Src-like adapter protein-2"
                     /protein_id="AAL29204.1"
                     /db_xref="GI:16797892"
                     /translation="MGSLPSRRKSLPSPSLSSSVQGQGPVTMEAERSKATAVALGSFP
                     AGGPAELSLRLGEPLTIVSEDGDWWTVLSEVSGREYNIPSVHVAKVSHGWLYEGLSRE
                     KAEELLLLPGNPGGAFLIRESQTRRGSYSLSVRLSRPASWDRIRHYRIHCLDNGWLYI
                     SPRLTFPSLQALVDHYSELADDICCLLKEPCVLQRAGPLPGKDIPLPVTVQRTPLNWK
                     ELDSSLLFSEAATGEESLLSEGLRESLSFYISLNDEAVSLDDA"
ORIGIN

Alignment Scores:
Pred. No.:                1.37e-101      Length:          786
Score:                    1351.00        Matches:         261
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       9              Gaps:            0

US-10-043-649-2 (1-261) x AF326353 (1-786)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660

Qy        221 AspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGly 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 720

Qy        241 LeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAspAsp 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 780

Qy        261 Ala 261
              |||
Db        781 GCC 783

RESULT 5
AK025645
LOCUS       AK025645               2415 bp    mRNA    linear   PRI 13-SEP-2003
DEFINITION  Homo sapiens cDNA: FLJ21992 fis, clone HEP06554.
ACCESSION   AK025645
VERSION     AK025645.1  GI:10438227
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
            Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
            Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2415)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Laboratory of Genome Structure Analysis, Human
            Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
            Japan (E-mail:flcdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Departent of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1. .2415
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HEP06554"
                     /cell_line="HepG2"
                     /cell_type="hepatoma"
                     /clone_lib="HEP"
                     /note="cloning vector pME18SFL3"
     CDS             61. .846
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB15201.1"
                     /db_xref="GI:10438228"
                     /translation="MGSLPSRRKSLPSPSLSSSVQGQGPVTMEAERSKATAVALGSFP
                     AGGPAELSLRLGEPLTIVSEDGDWWTVLSEVSGREYNIPSVHVAKVSHGWLYEGLSRE
                     KAEELLLLPGNPGGAFLIRESQTRRGSYSLSVRLSRPASWDRIRHYRIHCLDNGWLYI
                     SPRLTFPSLQALVDHYSELADDICCLLKEPCVLQRAGPLPGKDIPLPVTVQRTPLNWK
```

ELDSSLLFSEAATGEESLLSEGLRESLSFYISLNDEAVSLDDA"

ORIGIN

Alignment Scores:

Pred. No.:	5.04e-101	Length:	2415
Score:	1351.00	Matches:	261
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-10-043-649-2 (1-261) x AK025645 (1-2415)

```
Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 120

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 180

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 240

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 300

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 360

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 420

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 480

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 540

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 600

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 660

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 720

Qy        221 AspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGly 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 780

Qy        241 LeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAspAsp 260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 840

Qy        261 Ala 261
              |||
Db        841 GCC 843
```

RESULT 6
BC042041
LOCUS BC042041 2538 bp mRNA linear PRI 30-JUN-2004
DEFINITION Homo sapiens Src-like-adaptor 2, transcript variant 1, mRNA (cDNA clone MGC:49845 IMAGE:4429896), complete cds.
ACCESSION BC042041
VERSION BC042041.1 GI:27469842
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2538)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED 12477932
REFERENCE 2 (bases 1 to 2538)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (23-DEC-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 88 Row: a Column: 20
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 28416422.

FEATURES Location/Qualifiers

source	1. .2538
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="MGC:49845 IMAGE:4429896"
	/tissue_type="Prostate, adenocarcinoma."
	/clone_lib="NIH_MGC_91"
	/lab_host="DH10B"
	/note="Vector: pCMV-SPORT6"
gene	1. .2538
	/gene="SLA2"
	/note="synonyms: SLAP-2, MGC49845, FLJ21992, SLAP2, C20orf156"
	/db_xref="LocusID:84174"
	/db_xref="MIM:606577"
CDS	363. .1148

```
                     /gene="SLA2"
                     /codon_start=1
                     /product="Src-like-adaptor 2, isoform a"
                     /protein_id="AAH42041.1"
                     /db_xref="GI:27469843"
                     /db_xref="LocusID:84174"
                     /db_xref="MIM:606577"
                     /translation="MGSLPSRRKSLPSPSLSSSVQGQGPVTMEAERSKATAVALGSFP
                     AGGPAELSLRLGEPLTIVSEDGDWWTVLSEVSGREYNIPSVHVAKVSHGWLYEGLSRE
                     KAEELLLLPGNPGGAFLIRESQTRRGSYSLSVRLSRPASWDRIRHYRIHCLDNGWLYI
                     SPRLTFPSLQALVDHYSELADDICCLLKEPCVLQRAGPLPGKDIPLPVTVQRTPLNWK
                     ELDSSLLFSEAATGEESLLSEGLRESLSFYISLNDEAVSLDDA"
ORIGIN

Alignment Scores:
Pred. No.:                5.34e-101      Length:          2538
Score:                    1351.00        Matches:         261
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       9              Gaps:            0

US-10-043-649-2 (1-261) x BC042041 (1-2538)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 422

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 482

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 542

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 602

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 662

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 722

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        723 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 782

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        783 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 842

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        843 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 902

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        903 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 962

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        963 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 1022

Qy        221 AspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGly 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1023 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 1082

Qy        241 LeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAspAsp 260
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1083 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 1142

Qy        261 Ala 261
              |||
Db       1143 GCC 1145

RESULT 7
AX452880
LOCUS       AX452880                2567 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 1 from Patent WO0242457.
ACCESSION   AX452880
VERSION     AX452880.1  GI:21712520
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Chang,H., Yang,W.P., Wu,Y., Whitney,G.S., Perez-Villar,J.J. and
            Kanner,S.B.
  TITLE     Cloning and expression of human slap-2: a novel sh2/sh3
            domain-containing human slap homologue having immune cell-specific
            expression
  JOURNAL   Patent: WO 0242457-A 1 30-MAY-2002;
            Bristol-Myers Squibb Co. (US)
FEATURES             Location/Qualifiers
     source          1. .2567
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                5.41e-101      Length:          2567
Score:                    1351.00        Matches:         261
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       6              Gaps:            0

US-10-043-649-2 (1-261) x AX452880 (1-2567)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        415 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 474

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        475 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 534

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        535 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 594

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        595 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 654

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        655 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 714

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        715 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 774

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        775 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 834

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        835 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 894

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        895 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 954

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        955 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 1014

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1015 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 1074

Qy        221 AspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGly 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1075 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 1134

Qy        241 LeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAspAsp 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1135 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 1194

Qy        261 Ala 261
              |||
Db       1195 GCC 1197

RESULT 8
AX443133
LOCUS       AX443133               1183 bp    DNA     linear   PAT 02-JUL-2002
DEFINITION  Sequence 74 from Patent WO0216599.
ACCESSION   AX443133
VERSION     AX443133.1  GI:21690555
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Burgess,C.E., Conley,P.B., Grosse,W.M., Hart,M., Kekuda,R.,
            Shimkets,R.A., Spytek,K.A., Szekeres,E.S., Tomlinson,J.E.,
            Topper,J.N. and Yang,R.B.
  TITLE     Proteins and nucleic acids encoding same
  JOURNAL   Patent: WO 0216599-A 74 28-FEB-2002;
            Curagen Corporation (US) ; COR THERAPEUTICS, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .1183
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                4.68e-101      Length:          1183
Score:                    1347.00        Matches:         260
Percent Similarity:       99.62%         Conservative:    0
Best Local Similarity:    99.62%         Mismatches:      1
Query Match:              99.70%         Indels:          0
DB:                       6              Gaps:            0

US-10-043-649-2 (1-261) x AX443133 (1-1183)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        398 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 457

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        458 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 517

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        518 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 577

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        578 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 637

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ||||||||||||||||||||    ||||||||||||||||||||||||||||||||||||
Db        638 AACATCCCCAGCGTCCACGTGGGCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 697

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        698 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 757

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        758 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 817

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        818 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 877

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        878 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 937

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        938 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 997

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        998 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 1057

Qy        221 AspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGly 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1058 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 1117

Qy        241 LeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAspAsp 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1118 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 1177

Qy        261 Ala 261
              |||
Db       1178 GCC 1180

RESULT 9
AX443135/c
LOCUS       AX443135               1183 bp    DNA     linear   PAT 02-JUL-2002
DEFINITION  Sequence 76 from Patent WO0216599.
ACCESSION   AX443135
VERSION     AX443135.1  GI:21690556
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Burgess,C.E., Conley,P.B., Grosse,W.M., Hart,M., Kekuda,R.,
            Shimkets,R.A., Spytek,K.A., Szekeres,E.S., Tomlinson,J.E.,
            Topper,J.N. and Yang,R.B.
  TITLE     Proteins and nucleic acids encoding same
  JOURNAL   Patent: WO 0216599-A 76 28-FEB-2002;
            Curagen Corporation (US) ; COR THERAPEUTICS, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .1183
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                4.68e-101      Length:          1183
Score:                    1347.00        Matches:         260
Percent Similarity:       99.62%         Conservative:    0
```

```
Best Local Similarity:  99.62%        Mismatches:    1
Query Match:            99.70%        Indels:        0
DB:                     6             Gaps:          0

US-10-043-649-2 (1-261) x AX443135 (1-1183)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerServal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        786 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 727

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 667

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 607

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 547

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              |||||||||||||||||||||    ||||||||||||||||||||||||||||||||||||
Db        546 AACATCCCCAGCGTCCACGTGGGCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 487

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 427

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 367

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 307

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 247

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 187

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 127

Qy        221 AspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGly 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 67

Qy        241 LeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAspAsp 260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 7

Qy        261 Ala 261
              |||
Db          6 GCC 4

RESULT 10
AX780857
LOCUS       AX780857               2788 bp    DNA     linear   PAT 14-JUL-2003
DEFINITION  Sequence 3014 from Patent WO03039443.
ACCESSION   AX780857
VERSION     AX780857.1  GI:32697851
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1
  AUTHORS   Haferlach,T., Schoch,C., Kern,W., Kohlmann,A., Schnittger,S.,
            Dugas,M., Eils,R., Brors,B. and Mergenthaler,S.
  TITLE     Novel genetic markers for leukemias
  JOURNAL   Patent: WO 03039443-A 3014 15-MAY-2003;
            Deutsches Krebsforschungszentrum (DE) ;
            Ludwig-Maximilian-Universitaet Muenchen (DE) ; Haferlach, Torsten,
            PD Dr. Dr. (DE) ; Schoch, Claudia (DE) ; Kern, Wolfgang (DE)
FEATURES             Location/Qualifiers
     source          1. .2788
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              1.85e-100     Length:         2788
Score:                  1345.00       Matches:        260
Percent Similarity:     99.62%        Conservative:   0
Best Local Similarity:  99.62%        Mismatches:     1
Query Match:            99.56%        Indels:         0
DB:                     6             Gaps:           0

US-10-043-649-2 (1-261) x AX780857 (1-2788)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerServal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 446

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              ||||||    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 CAAGGCCANGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 506

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 566

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        567 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 626

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        627 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 686

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        687 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 746

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        747 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 806

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        807 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 866

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        867 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 926

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        927 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 986

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        987 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 1046

Qy        221 AspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGly 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1047 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGT 1106
```

```
Qy        241 LeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAspAsp 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1107 CTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGAT 1166

Qy        261 Ala 261
              |||
Db       1167 GCC 1169
```

RESULT 11
AX511155
LOCUS AX511155 737 bp DNA linear PAT 27-SEP-2002
DEFINITION Sequence 6 from Patent WO0242452.
ACCESSION AX511155
VERSION AX511155.1 GI:23392047
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Mcglade,J.C. and Loreto,M.P.
TITLE Adapter gene
JOURNAL Patent: WO 0242452-A 6 30-MAY-2002;
The Hospital for Sick Children (CA)
FEATURES Location/Qualifiers
source 1. .737
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN

Alignment Scores:

Pred. No.:	2.71e-89	Length:	737
Score:	1200.50	Matches:	241
Percent Similarity:	92.40%	Conservative:	2
Best Local Similarity:	91.63%	Mismatches:	1
Query Match:	88.86%	Indels:	19
DB:	6	Gaps:	1

US-10-043-649-2 (1-261) x AX511155 (1-737)

```
Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTAC---------- 528

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGln-ArgAlaGlyProLeuPr 200
                                                      ||||||||||||||||||
Db        529 ----------------------------------------TCTGAGGGCTGGCCCGCTCCC 549

Qy        200 oGlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLe 220
              ||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db        550 TGGCAAGGATATACCCCTACCTGTGACTGTGCGGAGGACACCACTCAACTGGAAAGAGCT 609

Qy        220 uAspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGl 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        610 GGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGG 669

Qy        240 yLeuArgGluSerLeuSerPheTyrIleSerLeuAsnAsp-GluAlaValSerLeuAspA 260
              ||||||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||
Db        670 TCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGAGCGAGGCTGTCTCTTTGGATG 729

Qy        260 spAla 261
              |||||
Db        730 ATGCC 734

RESULT 12
AF290986
LOCUS AF290986 737 bp mRNA linear PRI 21-JAN-2003
DEFINITION Homo sapiens Src-like adaptor protein-2 splice isoform mRNA,
complete cds; alternatively spliced.
ACCESSION AF290986
VERSION AF290986.1 GI:17351922
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 737)
AUTHORS Loreto,M.P. and McGlade,C.J.
TITLE Cloning and characterization of human Src-like adaptor protein 2
and a novel splice isoform, SLAP-2-v
JOURNAL Oncogene 22 (2), 266-273 (2003)
MEDLINE 22415750
PUBMED 12527895
REFERENCE 2 (bases 1 to 737)
AUTHORS Loreto,M.P. and McGlade,C.J.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-2000) Brain Tumour Research Centre, Hospital for
Sick Children, 555 University Avenue, Toronto, Ont M5G 1X8, Canada
FEATURES Location/Qualifiers
source 1. .737
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="20"
/tissue_type="thymus"
CDS 1. .633
/note="SLAP-2-v"
/codon_start=1
/product="Src-like adaptor protein-2 splice isoform"
/protein_id="AAL38198.1"
/db_xref="GI:17351923"
/translation="MGSLPSRRKSLPSPSLSSSVQGQGPVTMEAERSKATAVALGSFP
AGGPAELSLRLGEPLTIVSEDGDWWTVLSEVSGREYNIPSVHVAKVSHGWLYEGLSRE
KAEELLLLPGNPGGAFLIRESQTRRGSYSLSVRLSRPASWDRIRHYRIHCLDNGWLYI
SPRLTFPSLQALVDHYSEGWPAPWQGYTPTCDCAEDTTQLERAGQLPPVF"
ORIGIN

Alignment Scores:

| | | | |
|---|---|---|---|
| Pred. No.: | 2.71e-89 | Length: | 737 |
| Score: | 1200.50 | Matches: | 241 |
| Percent Similarity: | 92.40% | Conservative: | 2 |
| Best Local Similarity: | 91.63% | Mismatches: | 1 |
| Query Match: | 88.86% | Indels: | 19 |
```

```
DB:                    9             Gaps:          1

US-10-043-649-2 (1-261) x AF290986 (1-737)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              ||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTAC------------ 528

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGln-ArgAlaGlyProLeuPr 200
                                                       ||||||||||||||||||
Db        529 ------------------------------------TCTGAGGGCTGGCCCGCTCCC 549

Qy        200 oGlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLe 220
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db        550 TGGCAAGGATATACCCCTACCTGTGACTGTGCGGAGGACACCACTCAACTGGAAAGAGCT 609

Qy        220 uAspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGl 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        610 GGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGG 669

Qy        240 yLeuArgGluSerLeuSerPheTyrIleSerLeuAsnAsp-GluAlaValSerLeuAspA 260
              ||||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||||
Db        670 TCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGAGCGAGGCTGTCTCTTTGGATG 729

Qy        260 spAla 261
              |||||
Db        730 ATGCC 734

RESULT 13
AX511151
LOCUS       AX511151                 777 bp    DNA     linear   PAT 27-SEP-2002
DEFINITION  Sequence 2 from Patent WO0242452.
ACCESSION   AX511151
VERSION     AX511151.1  GI:23392045
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Mcglade,J.C. and Loreto,M.P.
```

```
  TITLE     Adapter gene
  JOURNAL   Patent: WO 0242452-A 2 30-MAY-2002;
            The Hospital for Sick Children (CA)
FEATURES             Location/Qualifiers
     source          1. .777
                     /organism="Mus musculus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10090"
ORIGIN

Alignment Scores:
Pred. No.:                 1.83e-75        Length:          777
Score:                     1032.00         Matches:         209
Percent Similarity:        85.88%          Conservative:    16
Best Local Similarity:     79.77%          Mismatches:      33
Query Match:               76.39%          Indels:          4
DB:                        6               Gaps:            3

US-10-043-649-2 (1-261) x AX511151 (1-777)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              ||||||||||||   ||||||   |||:::     |||||||||  |||||||||
Db          1 ATGGGAAGTTTGTCCAGCAGAGGGAAAACC---TCCAGCCCCAGCCCCAGCTCCTCTGGT 57

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
                 |||   ||||||:::|||:::   ||||||   |||   ||||||||||||||||
Db         58 CCAGACCAGGAACCCGTGTCCATGCAACCAGAAAGACACAAGGTCACAGCTGTGGCCCTG 117

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||   |||   ||||||||||||||||||||||||||||||||
Db        118 GGCAGTTTCCCAGCAGGTGAACAGGCCAGACTATCTCTGAGACTCGGGGAGCCGCTGACC 177

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        178 ATCATCTCTGAGGATGGAGATTGGTGGACAGTCCAGTCGGAAGTCTCAGGCAGAGAGTAC 237

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ::::::|||||||||:::|||||||||||||:::||||||||||||||||||||||||||
Db        238 CACATGCCCAGTGTGTATGTGGCTAAAGTCGCCCACGGGTGGCTGTACGAGGGCCTGAGC 297

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        298 CGGGAGAAAGCCGAGGAACTACTCCTGTTACCTGGGAACCCCGGAGGGGCCTTCCTCATC 357

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db        358 CGGGAGAGCCAGACCAGGAGAGGCTGCTATTCCCTGTCCGTCCGACTCAGCCGCCCTGCA 417

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              ||||||||||||||||||||||||      ||||||||||||||||||||||||||||||
Db        418 TCTTGGGACCGGATCAGACACTACAGGATACAGCGTCTTGACAATGGCTGGCTGTACATC 477

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              ||||||||||||||||||||||||   |||||||||:::||||||||||||||||||||
Db        478 TCACCTCGCCTCACCTTCCCCTCACTCCACGCCTTGGTGGAGCATTACTCTGAGCTAGCA 537

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              |||   |||||||||   |||:::||||||||||||||||||:::   ||||||||||||
Db        538 GATGGCATCTGCTGTCCCCTCAGGGAGCCGTGTGTCCTGCAGAAGCTTGGGCCACTACCT 597

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              |||||||||   |||   ||||||||||||      :::   ||||||||||||:::|||
Db        598 GGCAAAGATACACCTCCACCTGTGACTGTGCCAACATCATCACTAAATTGGAAAAAGCTG 657

Qy        221 AspSerSerLeuLeuPheSerGluAla---AlaThrGlyGluGluSerLeuLeuSerGlu 239
              |||   ||||||||||||   ||||||   |||:::|||||   ||||||||||||||||
Db        658 GACCGCAGCCTCCTGTTTCTGGAAGCACCTGCGAGTGGGGAGGCATCTCTGCTCAGTGAG 717

Qy        240 GlyLeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAsp 259
              ||||||||||||||||||||   ||||||||||||   ::::::      ||||||
Db        718 GGGCTCCGAGAGTCCCTCAGTTCCTACATCAGCCTGGCTGAGGAC------CCCTTGGAT 771

Qy        260 AspAla 261
```

```
                ||||||
Db        772 GATGCT 777

RESULT 14
AX511150
LOCUS       AX511150                1348 bp    DNA     linear   PAT 27-SEP-2002
DEFINITION  Sequence 1 from Patent WO0242452.
ACCESSION   AX511150
VERSION     AX511150.1  GI:23392044
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Mcglade,J.C. and Loreto,M.P.
  TITLE     Adapter gene
  JOURNAL   Patent: WO 0242452-A 1 30-MAY-2002;
            The Hospital for Sick Children (CA)
FEATURES             Location/Qualifiers
     source          1. .1348
                     /organism="Mus musculus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10090"
ORIGIN

Alignment Scores:
Pred. No.:                3.47e-75        Length:          1348
Score:                    1032.00         Matches:         209
Percent Similarity:       85.88%          Conservative:    16
Best Local Similarity:    79.77%          Mismatches:      33
Query Match:              76.39%          Indels:          4
DB:                       6               Gaps:            3

US-10-043-649-2 (1-261) x AX511150 (1-1348)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              |||||||||||   ||||||   |||:::      |||||||||   |||||||||
Db        282 ATGGGAAGTTTGTCCAGCAGAGGGAAAACC---TCCAGCCCCAGCCCCAGCTCCTCTGGT 338

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
                 |||   ||||||:::|||:::   ||||||   |||   ||||||||||||||||
Db        339 CCAGACCAGGAACCCGTGTCCATGCAACCAGAAAGACACAAGGTCACAGCTGTGGCCCTG 398

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||      |||   ||||||||||||||||||||||||||||||
Db        399 GGCAGTTTCCCAGCAGGTGAACAGGCCAGACTATCTCTGAGACTCGGGGAGCCGCTGACC 458

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              |||:::||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db        459 ATCATCTCTGAGGATGGAGATTGGTGGACAGTCCAGTCGGAAGTCTCAGGCAGAGAGTAC 518

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ::::::|||||||||:::|||||||||||||:::||||||||||||||||||||||||||
Db        519 CACATGCCCAGTGTGTATGTGGCTAAAGTCGCCCACGGGTGGCTGTACGAGGGCCTGAGC 578

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        579 CGGGAGAAAGCCGAGGAACTACTCCTGTTACCTGGGAACCCCGGAGGGGCCTTCCTCATC 638

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        639 CGGGAGAGCCAGACCAGGAGAGGCTGCTATTCCCTGTCCGTCCGACTCAGCCGCCCTGCA 698

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              |||||||||||||||||||||||||||||      ||||||||||||||||||||||||
Db        699 TCTTGGGACCGGATCAGACACTACAGGATACAGCGTCTTGACAATGGCTGGCTGTACATC 758

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              |||||||||||||||||||||||||||   |||||||||:::|||||||||||||||||||
Db        759 TCACCTCGCCTCACCTTCCCCTCACTCCACGCCTTGGTGGAGCATTACTCTGAGCTAGCA 818

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
```

```
              |||   |||||||||   |||:::||||||||||||||||||:::   ||||||||||||
Db        819 GATGGCATCTGCTGTCCCCTCAGGGAGCCGTGTGTCCTGCAGAAGCTTGGGCCACTACCT 878

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              |||||||||   |||   ||||||||||||      :::   |||||||||||||:::|||
Db        879 GGCAAAGATACACCTCCACCTGTGACTGTGCCAACATCATCACTAAATTGGAAAAAGCTG 938

Qy        221 AspSerSerLeuLeuPheSerGluAla---AlaThrGlyGluGluSerLeuLeuSerGlu 239
              |||   ||||||||||||   ||||||   |||:::||||||   ||||||||||||||||
Db        939 GACCGCAGCCTCCTGTTTCTGGAAGCACCTGCGAGTGGGGAGGCATCTCTGCTCAGTGAG 998

Qy        240 GlyLeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAsp 259
              ||||||||||||||||||||||   ||||||||||||   ::::::          ||||||
Db        999 GGGCTCCGAGAGTCCCTCAGTTCCTACATCAGCCTGGCTGAGGAC------CCCTTGGAT 1052

Qy        260 AspAla 261
              ||||||
Db       1053 GATGCT 1058

RESULT 15
AF287467
LOCUS       AF287467                1348 bp    mRNA    linear   ROD 03-JUN-2002
DEFINITION  Mus musculus Src-like adaptor protein-2 mRNA, complete cds.
ACCESSION   AF287467
VERSION     AF287467.1  GI:17351918
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1348)
  AUTHORS   Loreto,M.P., Berry,D.M. and McGlade,C.J.
  TITLE     Functional cooperation between c-Cbl and Src-like adaptor protein 2
            in the negative regulation of T-cell receptor signaling
  JOURNAL   Mol. Cell. Biol. 22 (12), 4241-4255 (2002)
  MEDLINE   22022020
   PUBMED   12024036
REFERENCE   2  (bases 1 to 1348)
  AUTHORS   Loreto,M.P. and McGlade,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2000) Brain Tumour Research Centre, Hospital for
            Sick Children, 555 University Avenue, Toronto, Ont M5G 1X8, Canada
FEATURES             Location/Qualifiers
     source          1. .1348
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /dev_stage="day 15 embryo"
     CDS             282. .1061
                     /note="SLAP-2"
                     /codon_start=1
                     /product="Src-like adaptor protein-2"
                     /protein_id="AAL38196.1"
                     /db_xref="GI:17351919"
                     /translation="MGSLSSRGKTSSPSPSSSGPDQEPVSMQPERHKVTAVALGSFPA
                     GEQARLSLRLGEPLTIISEDGDWWTVQSEVSGREYHMPSVYVAKVAHGWLYEGLSREK
                     AEELLLLPGNPGGAFLIRESQTRRGCYSLSVRLSRPASWDRIRHYRIQRLDNGWLYIS
                     PRLTFPSLHALVEHYSELADGICCPLREPCVLQKLGPLPGKDTPPPVTVPTSSLNWKK
                     LDRSLLFLEAPASGEASLLSEGLRESLSSYISLAEDPLDDA"
     misc_feature    381. .551
                     /note="Region: SH3 domain"
     misc_feature    552. .824
                     /note="Region: SH2 domain"
     polyA_signal    1301. .1306
ORIGIN

Alignment Scores:
Pred. No.:                3.47e-75        Length:          1348
Score:                    1032.00         Matches:         209
Percent Similarity:       85.88%          Conservative:    16
Best Local Similarity:    79.77%          Mismatches:      33
Query Match:              76.39%          Indels:          4
DB:                       10              Gaps:            3
```

US-10-043-649-2 (1-261) x AF287467 (1-1348)

```
Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              |||||||||||||    ||||||   |||:::     |||||||||   |||||||||
Db        282 ATGGGAAGTTTGTCCAGCAGAGGGAAAACC---TCCAGCCCCAGCCCCAGCTCCTCTGGT 338

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
                 |||   ||||||:::|||:::   ||||||   |||   ||||||||||||||||
Db        339 CCAGACCAGGAACCCGTGTCCATGCAACCAGAAAGACACAAGGTCACAGCTGTGGCCCTG 398

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||      |||   ||||||||||||||||||||||||||||
Db        399 GGCAGTTTCCCAGCAGGTGAACAGGCCAGACTATCTCTGAGACTCGGGGAGCCGCTGACC 458

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              |||:::|||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db        459 ATCATCTCTGAGGATGGAGATTGGTGGACAGTCCAGTCGGAAGTCTCAGGCAGAGAGTAC 518

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ::::::|||||||||:::|||||||||||||:::||||||||||||||||||||||||||
Db        519 CACATGCCCAGTGTGTATGTGGCTAAAGTCGCCCACGGGTGGCTGTACGAGGGCCTGAGC 578

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        579 CGGGAGAAAGCCGAGGAACTACTCCTGTTACCTGGGAACCCCGGAGGGGCCTTCCTCATC 638

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db        639 CGGGAGAGCCAGACCAGGAGAGGCTGCTATTCCCTGTCCGTCCGACTCAGCCGCCCTGCA 698

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              ||||||||||||||||||||||||||||      ||||||||||||||||||||||||||
Db        699 TCTTGGGACCGGATCAGACACTACAGGATACAGCGTCTTGACAATGGCTGGCTGTACATC 758

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              ||||||||||||||||||||||||   |||||||||:::|||||||||||||||||||
Db        759 TCACCTCGCCTCACCTTCCCCTCACTCCACGCCTTGGTGGAGCATTACTCTGAGCTAGCA 818

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              |||   |||||||||   |||:::|||||||||||||||||:::   ||||||||||||
Db        819 GATGGCATCTGCTGTCCCCTCAGGGAGCCGTGTGTCCTGCAGAAGCTTGGGCCACTACCT 878

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              |||||||||   |||   ||||||||||||      :::   ||||||||||||:::|||
Db        879 GGCAAAGATACACCTCCACCTGTGACTGTGCCAACATCATCACTAAATTGGAAAAAGCTG 938

Qy        221 AspSerSerLeuLeuPheSerGluAla---AlaThrGlyGluGluSerLeuLeuSerGlu 239
              |||   ||||||||||||   ||||||   |||:::||||||   ||||||||||||||
Db        939 GACCGCAGCCTCCTGTTTCTGGAAGCACCTGCGAGTGGGGAGGCATCTCTGCTCAGTGAG 998

Qy        240 GlyLeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAsp 259
              ||||||||||||||||||||   ||||||||||||   ::::::         ||||||
Db        999 GGGCTCCGAGAGTCCCTCAGTTCCTACATCAGCCTGGCTGAGGAC------CCCTTGGAT 1052

Qy        260 AspAla 261
              ||||||
Db       1053 GATGCT 1058
```

Search completed: November 17, 2004, 02:15:10
Job time : 4555 secs

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 17, 2004, 00:38:19 ; Search time 3371 Seconds
(without alignments)
2821.350 Million cell updates/sec

Title: US-10-043-649-2
Perfect score: 1351
Sequence: 1 MGSLPSRRKSLPSPSLSSSV.....RESLSFYISLNDEAVSLDDA 261

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10043649/runat_16112004_060536_28728/app_query.fasta_1.455
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10043649_@CGN_1_1_3437_@runat_16112004_060536_28728 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1277	94.5	1002	5	BQ052308	BQ052308 AGENCOURT
2	1187	87.9	1069	5	BQ052468	BQ052468 AGENCOURT
3	1152.5	85.3	1020	5	BQ054281	BQ054281 AGENCOURT
4	1032	76.4	2637	3	AK088672	AK088672 Mus muscu
5	1028	76.1	2974	3	AK030877	AK030877 Mus muscu
6	937.5	69.4	926	3	AK020837	AK020837 Mus muscu
7	882.5	65.3	660	6	BY742155	BY742155 BY742155
8	871	64.5	1032	1	AL541041	AL541041 AL541041
9	711	52.6	566	4	BG284179	BG284179 602408226
10	684.5	50.7	986	5	BQ054265	BQ054265 AGENCOURT
11	611.5	45.3	960	5	BU944126	BU944126 AGENCOURT
12	564.5	41.8	878	5	BQ053486	BQ053486 AGENCOURT
13	561	41.5	377	1	AA959151	AA959151 vz51g06.r
14	560.5	41.5	660	2	BB635615	BB635615 BB635615
15	555	41.1	597	1	AL844311	AL844311 AL844311
16	550	40.7	778	4	BG178487	BG178487 602328305
17	545	40.3	606	1	AL844309	AL844309 AL844309
18	543	40.2	614	1	AL844307	AL844307 AL844307
19	532	39.4	781	7	CK596391	CK596391 AGENCOURT
c 20	531	39.3	569	6	CB426333	CB426333 601508 MA
21	521.5	38.6	627	2	BB619854	BB619854 BB619854
22	520	38.5	762	5	BP162888	BP162888 BP162888
23	518	38.3	701	7	CK833360	CK833360 4057315 B
24	496	36.7	2810	3	AK036167	AK036167 Mus muscu
25	495.5	36.7	831	9	AY416279	AY416279 Mus muscu
26	495	36.6	1386	3	AK041565	AK041565 Mus muscu
27	493.5	36.5	831	9	AY416277	AY416277 Homo sapi
28	491	36.3	1762	3	CR617843	CR617843 full-leng
29	482	35.7	1037	1	AL539427	AL539427 AL539427
30	481.5	35.6	655	7	CF913437	CF913437 A0648D04-
31	480.5	35.6	1997	3	AK037901	AK037901 Mus muscu
32	477	35.3	966	1	AL551370	AL551370 AL551370
33	474.5	35.1	701	5	BX849096	BX849096 BX849096
34	472	34.9	1015	5	BX436423	BX436423 BX436423
35	471	34.9	972	5	BQ707614	BQ707614 AGENCOURT
36	468	34.6	1054	1	AL549826	AL549826 AL549826
37	461	34.1	656	7	CN792083	CN792083 4126904 B
38	460	34.0	796	7	CO574805	CO574805 AGENCOURT
39	456.5	33.8	871	5	BQ436143	BQ436143 AGENCOURT
40	451.5	33.4	775	6	CB938697	CB938697 IpCGJx13_
41	448.5	33.2	934	7	CK411626	CK411626 AUF_IpHdk
42	446	33.0	1063	5	BQ072745	BQ072745 AGENCOURT
43	445	32.9	723	5	BM950089	BM950089 UI-M-EH0p
44	443.5	32.8	777	6	CB938723	CB938723 IpCGJx13_
45	443.5	32.8	882	5	BX415149	BX415149 BX415149

ALIGNMENTS

RESULT 1
BQ052308
LOCUS BQ052308 1002 bp mRNA linear EST 29-MAR-2002
DEFINITION AGENCOURT_6868571 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:5933542 5', mRNA sequence.
ACCESSION BQ052308
VERSION BQ052308.1 GI:19811648
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1002)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2118 row: d column: 23
High quality sequence stop: 670.
FEATURES Location/Qualifiers
source 1. .1002
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5933542"

/tissue_type="natural killer cells, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_106"
/note="Organ: blood; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.:	1.89e-116	Length:	1002
Score:	1277.00	Matches:	257
Percent Similarity:	97.74%	Conservative:	2
Best Local Similarity:	96.98%	Mismatches:	2
Query Match:	94.52%	Indels:	4
DB:	5	Gaps:	0

US-10-043-649-2 (1-261) x BQ052308 (1-1002)

```
Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerServal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         96 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 155

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        156 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 215

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        216 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 275

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        276 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 335

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        336 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 395

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 455

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        456 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 515

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              |||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        516 TCCTGGGACCGGATCAGACACTACAGGATCTACTGCCTTGACAATGGCTGGCTGTACATC 575

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        576 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 635

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        636 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 695

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        696 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 755

Qy        221 AspSerSerLeuLeuPheSerGluAlaAlaThr-GlyGluGluSerLeuLeuSerGlu-G 240
              |||||||||||||||||||||||  ||||||  ||||||||||||||||||||||||| |
Db        756 GACAGCTCCCTCCTGTTTTCTGAAACTGCCACAGGGGGAGGAGTCTCTTCTCAGTGAGGG 815

Qy        240 lyLeuArg-GluSerLeuSerPheTyrIleSerLeuAsnAspGlu-AlaValSerLeuAs 259
              ||||||| |||||||||||||||||||||||   :::|||||| ||||||||||||||
```

```
Db        816 GTCTCCGGGGAGTCCCTCAGCTTCTACATCAGCCCTGATGACGAGGGCTGTCTCTTTGGA 875

Qy        259 pAspAla 261
                  |||||||
Db        876 TGATGCC 882
```

RESULT 2
BQ052468

```
LOCUS       BQ052468                1069 bp    mRNA    linear   EST 29-MAR-2002
DEFINITION  AGENCOURT_6868422 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:5933772
            5', mRNA sequence.
ACCESSION   BQ052468
VERSION     BQ052468.1  GI:19811808
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1069)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2118  row: n  column: 13
            High quality sequence stop: 681.
FEATURES             Location/Qualifiers
     source          1. .1069
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5933772"
                     /tissue_type="natural killer cells, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_106"
                     /note="Organ: blood; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN
```

Alignment Scores:

Pred. No.:	1.93e-107	Length:	1069
Score:	1187.00	Matches:	244
Percent Similarity:	92.42%	Conservative:	0
Best Local Similarity:	92.42%	Mismatches:	1
Query Match:	87.86%	Indels:	20
DB:	5	Gaps:	1

US-10-043-649-2 (1-261) x BQ052468 (1-1069)

```
Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerServal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         89 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 148

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        149 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 208

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        209 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 268
```

```
Qy        61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       269 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 328

Qy        81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       329 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 388

Qy       101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       389 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 448

Qy       121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       449 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 508

Qy       141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       509 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 568

Qy       161 SerProArgLeuThrPheProSerLeuGlnAlaLeu-ValAspHisTyrSerGluLeuAl 180
             |||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       569 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGGTGGACCATTACTCTGAGGG--- 625

Qy       180 aAspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPr 200
                                                          ||||||||||||||
Db       626 ----------------------------------------------GCTGGCCCGCTCCC 639

Qy       200 oGlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLe 220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       640 TGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCT 699

Qy       220 uAspSerSerLeuLeuPheSerGluAlaAlaThr-GlyGluGluSerLeuLeuSerGluG 240
             |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db       700 GGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGGAGGAGTCTCTTCTCAGTGAGG 759

Qy       240 lyLeuArgGlu-SerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAsp 259
             |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       760 GTCTCCGGGAAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGAT 819

Qy       260 AspAla 261
             ||||||
Db       820 GATGCC 825
```

```
RESULT 3
BQ054281
LOCUS       BQ054281                1020 bp    mRNA    linear   EST 29-MAR-2002
DEFINITION  AGENCOURT_6830234 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:5936362
            5', mRNA sequence.
ACCESSION   BQ054281
VERSION     BQ054281.1  GI:19813621
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1020)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2125  row: j  column: 11
            High quality sequence stop: 556.
FEATURES             Location/Qualifiers
     source          1. .1020
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5936362"
                     /tissue_type="natural killer cells, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_106"
                     /note="Organ: blood; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN
```

```
Alignment Scores:
Pred. No.:                4.93e-104      Length:          1020
Score:                    1152.50        Matches:         238
Percent Similarity:       93.51%         Conservative:    7
Best Local Similarity:    90.84%         Mismatches:      10
Query Match:              85.31%         Indels:          7
DB:                       5              Gaps:            1

US-10-043-649-2 (1-261) x BQ054281 (1-1020)

Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerServal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        212 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 271

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        272 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 331

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        332 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 391

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        392 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 451

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        452 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 511

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||    ||||||||||||||||||||||||
Db        512 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGNGAACCCTGGAGGGGCCTTCCTCATC 571

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        572 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 631

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        632 TCCTGNGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 691

Qy        161 SerProArgLeuThrPheProSerLeuGlnAla-LeuValAspHisTyrSerGluLeuAl 180
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db        692 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCCTGGTGGACCATTACTCTGAGCTGGC 751

Qy        180 aAspAspIleCysCysLeu-LeuLysGluProCysValLeuGlnArgAlaGlyProLeuP 200
              |||||||||||||||||| |||::::::|||||||||||||||||||  |||||||
Db        752 GGATGACATCTGCTGCCTANCTCAGGGACCCCTGTGTCCTGCAAAGGGCTGCCCCGCTCC 811

Qy        200 roGlyLysAspIle-ProLeuProValThrValGlnArgThrProLeuAsnTrpLysGlu 219
              |||||||||||||| |||||||||||||||||||||||||||   :::|||:::|||
Db        812 CTGGCAAAGATATACCCCCTACCTGTGACTGTGCAAAGGACACCCACTCACTGGGAAGAG 871

Qy        220 LeuAspSerSerLeuLeu---PheSerGluAlaAlaThrGlyGluGluSerLeuLeuSer 238
```

```
                ||||||||||||||||||   |||   :::          ||||||||||||||||||
Db        872 CTGGACAGCTCCCTCCTGGTTTTCTTGAAGCTGCCCCAGGGGGAGGAGTCTCTTCTCATT 931

Qy        239 GluGly--LeuArgGluSerLeuSerPheTyrIleSerLeu-AsnAspGluAla 255
              ||||||     ||||||||||||||||||||||||:::||| |||||||:::|||
Db        932 GAGGGGTCTTCCGGGAGTCCCTCAGCTTCTACATCAACCTGTAATGACCAGGCT 985
```

RESULT 4
AK088672

LOCUS	AK088672 2637 bp mRNA linear HTC 03-APR-2004
DEFINITION	Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN full-length enriched library, clone:E430023D24 product:MODULATOR OF ANTIGEN RECEPTOR SIGNALING MARS, full insert sequence.
ACCESSION	AK088672
VERSION	AK088672.1 GI:26353729
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1
AUTHORS	Carninci,P. and Hayashizaki,Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
REFERENCE	6 (bases 1 to 2637)
AUTHORS	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge. Please visit our web site for further details. URL:http://genome.gsc.riken.jp/ URL:http://fantom.gsc.riken.jp/.

FEATURES	Location/Qualifiers
source	1. .2637 /organism="Mus musculus" /mol_type="mRNA" /strain="NOD" /db_xref="FANTOM_DB:E430023D24" /db_xref="taxon:10090" /clone="E430023D24" /cell_type="thymic cells" /tissue_type="thymus" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="2 days neonate"
CDS	358. .1137 /note="unnamed protein product; MODULATOR OF ANTIGEN RECEPTOR SIGNALING MARS (SPTR|AAL38196, evidence: FASTY, 100%ID, 100%length, match=777) putative" /codon_start=1 /protein_id="BAC40495.1" /db_xref="GI:26353730" /translation="MGSLSSRGKTSSPSPSSSGPDQEPVSMQPERHKVTAVALGSFPA GEQARLSLRLGEPLTIISEDGDWWTVQSEVSGREYHMPSVYVAKVAHGWLYEGLSREK AEELLLLPGNPGGAFLIRESQTRRGCYSLSVRLSRPASWDRIRHYRIQRLDNGWLYIS PRLTFPSLHALVEHYSELADGICCPLREPCVLQKLGPLPGKDTPPPVTVPTSSLNWKK LDRSLLFLEAPASGEASLLSEGLRESLSSYISLAEDPLDDA"

ORIGIN

Alignment Scores:			
Pred. No.:	2.02e-91	Length:	2637
Score:	1032.00	Matches:	209
Percent Similarity:	85.88%	Conservative:	16
Best Local Similarity:	79.77%	Mismatches:	33
Query Match:	76.39%	Indels:	4
DB:	3	Gaps:	3

US-10-043-649-2 (1-261) x AK088672 (1-2637)

```
Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              ||||||||||||   ||||||   |||:::     |||||||||   |||||||||
Db        358 ATGGGAAGTTTGTCCAGCAGAGGGAAAACC---TCCAGCCCCAGCCCCAGCTCCTCTGGT 414

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
                 |||   ||||||:::|||:::   ||||||   |||   ||||||||||||||||
Db        415 CCAGACCAGGAACCCGTGTCCATGCAACCAGAAAGACACAAGGTCACAGCTGTGGCCCTG 474

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||     |||   ||||||||||||||||||||||||||||||
Db        475 GGCAGTTTCCCAGCAGGTGAACAGGCCAGACTATCTCTGAGACTCGGGGAGCCGCTGACC 534

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              |||:::|||||||||||||||||||||||||||   |||||||||||||||||||||||
Db        535 ATCATCTCTGAGGATGGAGATTGGTGGACAGTCCAGTCGGAAGTCTCAGGCAGAGAGTAC 594

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ::::::|||||||||:::|||||||||||||:::|||||||||||||||||||||||||
```

```
Db        595 CACATGCCCAGTGTGTATGTGGCTAAAGTCGCCCACGGGTGGCTGTACGAGGGCCTGAGC 654

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120

Db        655 CGGGAGAAAGCCGAGGAACTACTCCTGTTACCTGGGAACCCCGGAGGGGCCTTCCTCATC 714

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140

Db        715 CGGGAGAGCCAGACCAGGAGAGGCTGCTATTCCCTGTCCGTCCGACTCAGCCGCCCTGCA 774

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160

Db        775 TCTTGGGACCGGATCAGACACTACAGGATACAGCGTCTTGACAATGGCTGGCTGTACATC 834

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180

Db        835 TCACCTCGCCTCACCTTCCCCTCACTCCACGCCTTGGTGGAGCATTACTCTGAGCTAGCA 894

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200

Db        895 GATGGCATCTGCTGTCCCCTCAGGGAGCCGTGTGTCCTGCAGAAGCTTGGGCCACTACCT 954

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220

Db        955 GGCAAAGATACACCTCCACCTGTGACTGTGCCAACATCATCACTAAATTGGAAAAAGCTG 1014

Qy        221 AspSerSerLeuLeuPheSerGluAla---AlaThrGlyGluGluSerLeuLeuSerGlu 239

Db       1015 GACCGCAGCCTCCTGTTTCTGGAAGCACCTGCGAGTGGGGAGGCATCTCTGCTCAGTGAG 1074

Qy        240 GlyLeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAsp 259

Db       1075 GGGCTCCGAGAGTCCCTCAGTTCCTACATCAGCCTGGCTGAGGAC------CCCTTGGAT 1128

Qy        260 AspAla 261

Db       1129 GATGCT 1134
```

RESULT 5
AK030877

LOCUS	AK030877 2974 bp mRNA linear HTC 03-APR-2004
DEFINITION	Mus musculus adult male thymus cDNA, RIKEN full-length enriched library, clone:5830437K10 product:MODULATOR OF ANTIGEN RECEPTOR SIGNALING MARS, full insert sequence.
ACCESSION	AK030877
VERSION	AK030877.1 GI:26326848
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1
AUTHORS	Carninci,P. and Hayashizaki,Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
REFERENCE	6 (bases 1 to 2974)
AUTHORS	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.jp/ URL:http://fantom.gsc.riken.jp/.

FEATURES Location/Qualifiers

```
     source          1. .2974
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:5830437K10"
                     /db_xref="taxon:10090"
                     /clone="5830437K10"
                     /sex="male"
                     /tissue_type="thymus"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             232. .1011
                     /note="unnamed protein product; MODULATOR OF ANTIGEN
                     RECEPTOR SIGNALING MARS (SPTR|AAL38196, evidence: FASTY,
                     100%ID, 100%length, match=777)
                     putative"
                     /codon_start=1
                     /protein_id="BAC27168.1"
                     /db_xref="GI:26326849"
                     /translation="MGSLSSRGKTSSPSPSSSGPDQEPVSMQPERHKVTAVALGSFPA
                     GEQARLSLRLGEPLTIISEDGDWWTVQSEVSGREYHMPSVYVAKVAHGWLYEGLSREK
                     AEELLLLPGNPGGAFLIRESQTRRGYYSLSVRLSRPASWDRIRHYRIQRLDNGWLYIT
                     PRLTFPSLHALVEHYSELADGICCPLREPCVLQKLGPLPGKDTPPPVTVPTSSLNWKK
                     LDRSLLFLEAPASGEASLLSEGLRESLSSYISLAEDPLDDA"
```

ORIGIN

Alignment Scores:

Pred. No.:	6.05e-91	Length:	2974
Score:	1028.00	Matches:	208
Percent Similarity:	85.88%	Conservative:	17
Best Local Similarity:	79.39%	Mismatches:	33
Query Match:	76.09%	Indels:	4
DB:	3	Gaps:	3

US-10-043-649-2 (1-261) x AK030877 (1-2974)

```
Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              |||||||||||||    ||||||   |||:::     |||||||||   |||||||||
Db        232 ATGGGAAGTTTGTCCAGCAGAGGGAAAACC---TCCAGCCCCAGCCCCAGCTCCTCTGGT 288

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
                 |||   ||||||:::|||:::   ||||||   |||   ||||||||||||||||
Db        289 CCAGACCAGGAACCCGTGTCCATGCAACCAGAAAGACACAAGGTCACAGCTGTGGCCCTG 348

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              |||||||||||||||||      |||   |||||||||||||||||||||||||||||
Db        349 GGCAGTTTCCCAGCAGGTGAACAGGCCAGACTATCTCTGAGACTCGGGGAGCCGCTGACC 408

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              |||:::|||||||||||||||||||||||||||   |||||||||||||||||||||||
Db        409 ATCATCTCTGAGGATGGAGATTGGTGGACAGTCCAGTCGGAAGTCTCAGGCAGAGAGTAC 468

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ::::::|||||||||:::|||||||||||||:::|||||||||||||||||||||||||
Db        469 CACATGCCCAGTGTGTATGTGGCTAAAGTCGCCCACGGGTGGCTGTACGAGGGCCTGAGC 528

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        529 CGGGAGAAAGCCGAGGAACTACTCCTGTTACCTGGGAACCCCGGAGGGGCCTTCCTCATC 588

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              |||||||||||||||||||||   |||||||||||||||||||||||||||||||||||
Db        589 CGGGAGAGCCAGACCAGGAGAGGCTACTATTCCCTGTCCGTCCGACTCAGCCGCCCTGCA 648

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              |||||||||||||||||||||||||||      |||||||||||||||||||||||||
Db        649 TCTTGGGACCGGATCAGACACTACAGGATACAGCGTCTTGACAATGGCTGGCTGTACATC 708

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
              :::|||||||||||||||||||||||   |||||||||:::|||||||||||||||||||
Db        709 ACACCTCGCCTCACCTTCCCCTCACTCCACGCCTTGGTGGAGCATTACTCTGAGCTAGCA 768

Qy        181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
              |||   |||||||||   |||:::|||||||||||||||||:::   ||||||||||||
Db        769 GATGGCATCTGCTGTCCCCTCAGGGAGCCGTGTGTCCTGCAGAAGCTTGGGCCACTACCT 828

Qy        201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
              |||||||||   |||   ||||||||||||      :::   ||||||||||||:::|||
Db        829 GGCAAAGATACACCTCCACCTGTGACTGTGCCAACATCATCACTAAATTGGAAAAAGCTG 888

Qy        221 AspSerSerLeuLeuPheSerGluAla---AlaThrGlyGluGluSerLeuLeuSerGlu 239
              |||   ||||||||||||   ||||||   |||:::|||||   ||||||||||||||||
Db        889 GACCGCAGCCTCCTGTTTCTGGAAGCACCTGCGAGTGGGGAGGCATCTCTGCTCAGTGAG 948

Qy        240 GlyLeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAsp 259
              |||||||||||||||||||   ||||||||||||   ::::::          ||||||
Db        949 GGGCTCCGAGAGTCCCTCAGTTCCTACATCAGCCTGGCTGAGGAC------CCCTTGGAT 1002

Qy        260 AspAla 261
              ||||||
Db       1003 GATGCT 1008
```

RESULT 6
AK020837

LOCUS	AK020837 926 bp mRNA linear HTC 03-APR-2004
DEFINITION	Mus musculus adult retina cDNA, RIKEN full-length enriched library, clone:A930009E21 product:MODULATOR OF ANTIGEN RECEPTOR SIGNALING MARS, full insert sequence.
ACCESSION	AK020837
VERSION	AK020837.1 GI:12861542
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1
AUTHORS	Carninci,P. and Hayashizaki,Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
REFERENCE	6 (bases 1 to 926)
AUTHORS	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	Please visit our web site (http://genome.gsc.riken.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase

and subsequently enriched for full-length by cap-trapper. cDNA went through two rounds of normalization to Rot = 20.0 and subtraction to Rot = 458.8. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B. Retina RNA was provided by Stefano Gustincich (Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA 02115, USA) whose assistance is gratefully acknowledged.
Retina RNA was provided by Stefano Gustincich (Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA 02115, USA) whose assistance is gratefully acknowledged.

```
FEATURES             Location/Qualifiers
     source          1. .926
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:A930009E21"
                     /db_xref="taxon:10090"
                     /clone="A930009E21"
                     /tissue_type="retina"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             128. .667
                     /note="unnamed protein product; MODULATOR OF ANTIGEN
                     RECEPTOR SIGNALING MARS (SPTR|AAL38196, evidence: FASTY,
                     100%ID, 100%length, match=777)
                     putative"
                     /codon_start=1
                     /protein_id="BAB32223.1"
                     /db_xref="GI:12861543"
                     /translation="MPSVYVAKVAHGWLYEGLSREKAEELLLLPGNPGGAFLIRESQT
                     RRGCYSLSVRLSRPASWDRIRHYRIQRLDNGWLYISPRLTFPSLHALVEHYSELADGI
                     CCPLREPCVLQKLGPLPGKDTPPPVTVPTSSLNWKKLDRSLLFLEAPASGEASLLSEG
                     LRESLSSYISLAEDPLDDA"
     polyA_signal    907. .912
                     /note="putative"
     polyA_site      926
                     /note="putative"
ORIGIN
```

Alignment Scores:

Pred. No.:	1.13e-82	Length:	926
Score:	937.50	Matches:	184
Percent Similarity:	88.74%	Conservative:	13
Best Local Similarity:	82.88%	Mismatches:	22
Query Match:	69.39%	Indels:	3
DB:	3	Gaps:	2

US-10-043-649-2 (1-261) x AK020837 (1-926)

```
Qy          41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
               ||||||||||||||||||      |||   ||||||||||||||||||||||||||||||
Db           5 GGCAGTTTCCCAGCAGGTGAACAGGCCAGACTATCTCTGAGACTCGGGGAGCCGCTGACC 64

Qy          61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
               |||:::||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db          65 ATCATCTCTGAGGATGGAGATTGGTGGACAGTCCAGTCGGAAGTCTCAGGCAGAGAGTAC 124

Qy          81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
               ::::::|||||||||:::|||||||||||||:::|||||||||||||||||||||||||||
Db         125 CACATGCCCAGTGTGTATGTGGCTAAAGTCGCCCACGGGTGGCTGTACGAGGGCCTGAGC 184

Qy         101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         185 CGGGAGAAAGCCGAGGAACTACTCCTGTTACCTGGGAACCCCGGAGGGGCCTTCCTCATC 244

Qy         121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
               ||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db         245 CGGGAGAGCCAGACCAGGAGAGGCTGCTATTCCCTGTCCGTCCGACTCAGCCGCCCTGCA 304

Qy         141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
               ||||||||||||||||||||||||||||||      ||||||||||||||||||||||||
Db         305 TCTTGGGACCGGATCAGACACTACAGGATACAGCGTCTTGACAATGGCTGGCTGTACATC 364

Qy         161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
               ||||||||||||||||||||||||||||   ||||||||:::||||||||||||||||||
Db         365 TCACCTCGCCTCACCTTCCCCTCACTCCACGCCTTGGTGGAGCATTACTCTGAGCTAGCA 424

Qy         181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
               |||   ||||||||||   |||:::|||||||||||||||||:::   ||||||||||||
Db         425 GATGGCATCTGCTGTCCCCTCAGGGAGCCGTGTGTCCTGCAGAAGCTTGGGCCACTACCT 484

Qy         201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
               |||||||||   |||   ||||||||||||      :::   ||||||||||||:::|||
Db         485 GGCAAAGATACACCTCCACCTGTGACTGTGCCAACATCATCACTAAATTGGAAAAAGCTG 544

Qy         221 AspSerSerLeuLeuPheSerGluAla---AlaThrGlyGluGluSerLeuLeuSerGlu 239
               |||   ||||||||||||   ||||||   |||:::||||||   ||||||||||||||||
Db         545 GACCGCAGCCTCCTGTTTCTGGAAGCACCTGCGAGTGGGGAGGCATCTCTGCTCAGTGAG 604

Qy         240 GlyLeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAsp 259
               ||||||||||||||||||||   ||||||||||||   ::::::          ||||||
Db         605 GGGCTCCGAGAGTCCCTCAGTTCCTACATCAGCCTGGCTGAGGAC------CCCTTGGAT 658

Qy         260 AspAla 261
               ||||||
Db         659 GATGCT 664
```

```
RESULT 7
BY742155
LOCUS       BY742155                660 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION  BY742155 RIKEN full-length enriched, adult retina Mus musculus cDNA
            clone A930009E21 5', mRNA sequence.
ACCESSION   BY742155
VERSION     BY742155.1  GI:27167493
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 660)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
```

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Retina RNA was provided by Dr. Stefano Gustincich (Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA 02115, USA) whose assistance is gratefully acknowledged.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .660
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A930009E21"
/tissue_type="retina"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult retina"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5'GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. ~Retina RNA was provided by Stefano Gustincich, Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA02115, USA, whose assistance we gratefully acknowledge."

ORIGIN

Alignment Scores:

Pred. No.:	2.07e-77	Length:	660
Score:	882.50	Matches:	175
Percent Similarity:	85.97%	Conservative:	15
Best Local Similarity:	79.19%	Mismatches:	24
Query Match:	65.32%	Indels:	8
DB:	6	Gaps:	2

US-10-043-649-2 (1-261) x BY742155 (1-660)

```
Qy        41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
             ||||||||||||||||||     |||   ||||||||||||||||||||||||||||||
Db         5 GGCAGTTTCCCAGCAGGTGAACAGGCCAGACTATCTCTGAGACTCGGGGAGCCGCTGACC 64

Qy        61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
             |||:::|||||||||||||||||||||||||||   |||||||||||||||||||||||
Db        65 ATCATCTCTGAGGATGGAGATTGGTGGACAGTCCAGTCGGAAGTCTCAGGCAGAGAGTAC 124

Qy        81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
             ::::::|||||||||:::|||||||||||||:::|||||||||||||||||||||||||
Db       125 CACATGCCCAGTGTGTATGTGGCTAAAGTCGCCCACGGGTGGCTGTACGAGGGCCTGAGC 184

Qy       101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       185 CGGGAGAAAGCCGAGGAACTACTCCTGTTACCTGGGAACCCCGGAGGGGCCTTCCTCATC 244

Qy       121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
             ||||||||||||||||||||||   |||||||||||||||||||||||||||||||||||
Db       245 CGGGAGAGCCAGACCAGGAGAGGCTGCTATTCCCTGTCCGTCCGACTCAGCCGCCCTGCA 304

Qy       141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
             ||||||||||||||||||||||||||||||      ||||||||||||||||||||||||
Db       305 TCTTGGGACCGGATCAGACACTACAGGATACAGCGTCTTGACAATGGCTGGCTGTACATC 364

Qy       161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAla 180
             ||||||||||||||||||||||||||||   |||||||||:::|||||||||||||||||
Db       365 TCACCTCGCCTCACCTTCCCCTCACTCCACGCCTTGGTGGAGCATTACTCTGAGCTAGCA 424

Qy       181 AspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuPro 200
             |||   |||||||||   |||:::|||||||||||||||||:::   ||||||||||||
Db       425 GATGGCATCTGCTGTCCCCTCAGGGAGCCGTGTGTCCTGCAGAAGCTTGGGCCACTACCT 484

Qy       201 GlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeu 220
             ||||||||   |||   ||||||||||||      :::   ||||||||||||:::|||
Db       485 GGCAAAGATACACCTCCACCTGTGACTGTGCCAACATCATCACTAAATTGGAAAAAGCTG 544

Qy       221 AspSerSerLeuLeuPheSerGluAla---AlaThrGlyGluGluSerLeuLeuSerGlu 239
             |||   ||||||||||||   ||||||   |||:::||||||   |||||||||   |||
Db       545 GACCGTAGCCTCCTGTTTCTGGAAGCACCTGCAAGTGGGGAGGCATCTCTGCTCATTGAG 604

Qy       240 GlyLeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAsp 259
             |||       |||   |||:::||||||:::                  :::   :::|||
Db       605 GGGCCTCCAGAGTC-CTCACCTTCTACCTC------------------ATCCCGGTTGAT 645

Qy       260 Asp 260
             |||
Db       646 GAC 648
```

RESULT 8
AL541041
LOCUS AL541041 1032 bp mRNA linear EST 24-MAR-2004
DEFINITION AL541041 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE005YK23 5-PRIME, mRNA sequence.
ACCESSION AL541041
VERSION AL541041.3 GI:45716635
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1032)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 15, 2001 this sequence version replaced gi:30544829.
Contact: Genoscope
Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized. Library was constructed by Life Technologies, a division of Invitrogen.
This sequence belongs to sequence cluster 9825.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0DE005AF12QP1&c=9825.r.

FEATURES Location/Qualifiers
source 1. .1032
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE005YK23"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Alignment Scores:

Pred. No.:	5.59e-76	Length:	1032
Score:	871.00	Matches:	169
Percent Similarity:	96.02%	Conservative:	0
Best Local Similarity:	96.02%	Mismatches:	7
Query Match:	64.47%	Indels:	0
DB:	1	Gaps:	0

US-10-043-649-2 (1-261) x AL541041 (1-1032)

```
Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        395 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 454

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        455 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 514

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        515 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 574

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        575 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 634

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        635 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 694

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        695 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 754

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        755 CGGGAGAGCCAGAMCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 814

Qy        141 SerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIle 160
              ||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db        815 TCCTGGGACCGGATCAGACACTAMAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 874

Qy        161 SerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyr 176
              |||   ||||||   |||   ||||||||||||||||||   |||
Db        875 TCACGGCGCCTYAACYTYCCCYCACTCCAGGCCCTGGTGGACMATTAC 922
```

RESULT 9
BG284179

LOCUS BG284179 566 bp mRNA linear EST 21-FEB-2001
DEFINITION 602408226F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4520382 5', mRNA sequence.
ACCESSION BG284179
VERSION BG284179.1 GI:13034866
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 566)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10418 row: c column: 07
High quality sequence start: 2
High quality sequence stop: 566.

FEATURES Location/Qualifiers
source 1. .566
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4520382"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_91"
/note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.4 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.:	2.02e-60	Length:	566
Score:	711.00	Matches:	136
Percent Similarity:	99.27%	Conservative:	0
Best Local Similarity:	99.27%	Mismatches:	1
Query Match:	52.63%	Indels:	0
DB:	4	Gaps:	0

US-10-043-649-2 (1-261) x BG284179 (1-566)

```
Qy        125 ThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAlaSerTrpAspArg 144
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 ACGCGTCCGGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCATCCTGGGACCGG 63

Qy        145 IleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyrIleSerProArgLeu 164
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 ATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATCTCACCGCGCCTC 123

Qy        165 ThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeuAlaAspAspIleCys 184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 ACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCGGATGACATCTGC 183

Qy        185 CysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLeuProGlyLysAspIle 204
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 TGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCTGGCAAGGATATA 243

Qy        205 ProLeuProValThrValGlnArgThrProLeuAsnTrpLysGluLeuAspSerSerLeu 224
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 CCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTGGACAGCTCCCTC 303

Qy        225 LeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGluGlyLeuArgGluSer 244
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 CTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGTCTCCGGGAGTCC 363

Qy        245 LeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAspAspAla 261
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 CTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGATGCC 414
```

RESULT 10
BQ054265

```
LOCUS       BQ054265                986 bp    mRNA    linear   EST 29-MAR-2002
DEFINITION  AGENCOURT_6830248 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:5936339
            5', mRNA sequence.
ACCESSION   BQ054265
VERSION     BQ054265.1  GI:19813605
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 986)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2125  row: i  column: 12
            High quality sequence stop: 515.
FEATURES             Location/Qualifiers
     source          1. .986
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5936339"
                     /tissue_type="natural killer cells, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_106"
                     /note="Organ: blood; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN
```

Alignment Scores:

Pred. No.:	2e-57	Length:	986
Score:	684.50	Matches:	164
Percent Similarity:	73.57%	Conservative:	3
Best Local Similarity:	72.25%	Mismatches:	16
Query Match:	50.67%	Indels:	44
DB:	5	Gaps:	3

US-10-043-649-2 (1-261) x BQ054265 (1-986)

```
Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 338

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 398

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        399 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 458

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        459 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 518

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        519 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 578

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAla-PheLeuIl 120
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        579 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCCTTCCTCAT 638

Qy        120 eArgGlu-SerGlnThrArgArgGlySerTyrSerLeuSerVal-ArgLeuSerArg-Pr 139
              ||||||| |||||||||||||||||||||||||||||||||| |||||||||||| ||
Db        639 CCGGGAGAAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCCGCCTCAGCCGCCCC 698

Qy        139 oAlaSerTrp-AspArg-IleArgHisTyrArg-IleHisCysLeuAspAsnGlyTrpLe 158
              |||||||||| |||||| |||||| |||||| ||| ||| |||||
Db        699 TGCATCCTGGGGACCGGGATCAGACCCTACAGGGATTCCCCTGCCCTTGAACCATTGGCT 758

Qy        158 u------------TyrIleSerProArgLeuThrPheProSerLeuGlnAlaLeuValAs 174
              |            :::   :::    |||   ||||||
Db        759 TGGCTTGTTAAATTTTTAAACCGGGGCCTTACCCTTTCCC------------------- 798

Qy        174 pHisTyrSerGluLeuAlaAspAspIleCysCysLeuLeuLysGluProCysValLeuGl 194
                                                                  |||::
Db        799 ----------------------------------------------------CTTAA 803

Qy        194 nArgAlaGlyProLeuProGlyLysAspIleProLeuProValThrValGlnArgThrPr 214
              :     |||||||||   |||                                 |||||
Db        804 ATTTCAGGGGCCCCTTGGGGGGAACCCATTT----------------------ACTCC 839

Qy        214 oLeuAsn 216
                |||||||
Db        840 TTTAAAC 846
```

RESULT 11
BU944126

```
LOCUS       BU944126                960 bp    mRNA    linear   EST 18-OCT-2002
DEFINITION  AGENCOURT_10545003 NIH_MGC_107 Homo sapiens cDNA clone
            IMAGE:6728350 5', mRNA sequence.
ACCESSION   BU944126
VERSION     BU944126.1  GI:24132945
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 960)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM3049  row: m  column: 21
            High quality sequence stop: 628.
FEATURES             Location/Qualifiers
     source          1. .960
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6728350"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
```

/clone_lib="NIH_MGC_107"
/note="Organ: breast; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:

Pred. No.:	3.61e-50	Length:	960
Score:	611.50	Matches:	124
Percent Similarity:	86.71%	Conservative:	0
Best Local Similarity:	86.71%	Mismatches:	1
Query Match:	45.26%	Indels:	18
DB:	5	Gaps:	1

US-10-043-649-2 (1-261) x BU944126 (1-960)

```
Qy        120 IleArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgPro 139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 ATCCGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCT 62

Qy        140 AlaSerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyr 159
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 GCATCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTAC 122

Qy        160 IleSerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLeu 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 ATCTCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTAC--------- 173

Qy        180 AlaAspAspIleCysCysLeuLeuLysGluProCysValLeuGln-ArgAlaGlyProLe 199
                                                          ||||||||||||||
Db        174 -----------------------------------------TCTGAGGGCTGGCCCGCT 191

Qy        199 uProGlyLysAspIleProLeuProValThrValGlnArgThrProLeuAsnTrpLysGl 219
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 CCCTGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGA 251

Qy        219 uLeuAspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGluSerLeuLeuSerGl 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 GCTGGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGA 311

Qy        239 uGlyLeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGluAlaValSerLeuAs 259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 GGGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGA 371

Qy        259 pAspAla 261
              |||||||
Db        372 TGATGCC 378
```

RESULT 12
BQ053486
LOCUS BQ053486 878 bp mRNA linear EST 29-MAR-2002
DEFINITION AGENCOURT_6822017 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:5935253 5', mRNA sequence.
ACCESSION BQ053486
VERSION BQ053486.1 GI:19812826
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 878)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2122 row: l column: 06
High quality sequence stop: 394.
FEATURES Location/Qualifiers
source 1. .878
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5935253"
/tissue_type="natural killer cells, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_106"
/note="Organ: blood; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:

Pred. No.:	1.52e-45	Length:	878
Score:	564.50	Matches:	135
Percent Similarity:	64.04%	Conservative:	11
Best Local Similarity:	59.21%	Mismatches:	20
Query Match:	41.78%	Indels:	64
DB:	5	Gaps:	5

US-10-043-649-2 (1-261) x BQ053486 (1-878)

```
Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerServal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        273 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 332

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        333 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 392

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        393 CGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 452

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              ||||||||||||||||||||||||||||||||||||:::|||   |||||||||
Db        453 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAATCTCACGCAGAGAGTAT 512

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              |||||||||   ||||||||||||||||||||||   ||||||||||||||
Db        513 AACATCCCCATCGTCCACGTGGCCAAAGTCTCCCATGCGTGGCTGTATGAGGGCT--GAN 570

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              :::||||||   :::   ||||||||||||||||||   ||||||||||||||||:::
Db        571 CAAGAAAAAACCAAAGGACTGCTGTTGTTACCTGGGAAACCTGGAGGGGCCTTTCTCCTC 630

Qy        121 -ArgGluSerGlnThrArgArgGlySer---TyrSerLeuSerValArgLeuSerArgPr 139
               ||||||:::|||||||||||||||||      |||:::  ||||||||||||||||||
Db        631 CCGGGAGAACCAGACCAGGAGAGGCTCCTAACTCTCTGTGTA-GTCCGCCTTAGCCGCCC 689

Qy        139 oAlaSerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTy 159
              |
Db        690 C---------------------------------------------------------- 690

Qy        159 rIleSerProArgLeuThrPheProSerLeuGlnAlaLeuValAspHisTyrSerGluLe 179

Db        690 ---------------------------------------------------------- 690

Qy        179 uAlaAspAspIleCysCysLeuLeuLysGluProCysValLeuGlnArgAlaGlyProLe 199
```

```
                          |||:::      ||||||             :::   |||
Db        691 ----------------TGCATTCCTGGGGAACCG-----------GATTCGAAACCTTA 722

Qy        199 uProGlyLysAspIleProLeuProValThr--------------------------Va 210
               ||||||          |||   |||:::|||                        ::
Db        723 CCCGGGACCC------CCCTGGCCTTTGACAAATGGGGCTTGGGCTTTGGAAAATTTTTT 776

Qy        210 lGlnArgThrProLeuAsnTrp 217
              :   |||   |||||||||:::
Db        777 ACACCGGGGGCCCTTAAACTTT 798
```

RESULT 13
AA959151
LOCUS AA959151 377 bp mRNA linear EST 08-MAY-1998
DEFINITION vz51g06.r1 Soares_thymus_2NbMT Mus musculus cDNA clone IMAGE:1330042 5' similar to TR:Q13239 Q13239 PUTATIVE SRC-LIKE ADAPTER PROTEIN ;, mRNA sequence.
ACCESSION AA959151
VERSION AA959151.1 GI:3124344
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 377)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:689586
Seq primer: -28m13 rev2 ET from Amersham.

FEATURES Location/Qualifiers
source 1. .377
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1330042"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/clone_lib="Soares_thymus_2NbMT"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN

Alignment Scores:

Pred. No.:	9.8e-46	Length:	377
Score:	561.00	Matches:	110
Percent Similarity:	93.50%	Conservative:	5
Best Local Similarity:	89.43%	Mismatches:	8
Query Match:	41.52%	Indels:	1

```
DB:                    1          Gaps:          0

US-10-043-649-2 (1-261) x AA959151 (1-377)

Qy         40 LeuGlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeu 59
              ||||||||||||||||||||||      |||   |||||||||||||||||||||||||
Db          8 CTGGGCAGTTTCCCAGCAGGTGAACAGGCCAGACTATCTCTGAGACTCGGGGAGCCGCTG 67

Qy         60 ThrIleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGlu 79
              ||||||:::||||||||||||||||||||||||||   ||||||||||||||||||||
Db         68 ACCATCATCTCTGAGGATGGAGATTGGTGGACAGTCCAGTCGGAAGTCTCAGGCAGAGAG 127

Qy         80 TyrAsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeu 99
              |||::::::|||||||||:::|||||||||||:::|||||||||||||||||||||||
Db        128 TACCACATGCCCAGTGTGTATGTGGCTAAAGTCGCCCACGGGTGGCTGTACGAGGGCCTG 187

Qy        100 SerArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeu 119
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 AGCCGGGAGAAAGCCGAGGAACTACTCCTGTTACCTGGGAACCCCGGAGGTCA-TTCCTC 246

Qy        120 IleArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgPro 139
              ||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db        247 ATCCGGGAGAGCCAGACCAGGAGAGGCTGCTATTCCCTGTCCGTCCGACTCAGCCGCCCT 306

Qy        140 AlaSerTrpAspArgIleArgHisTyrArgIleHisCysLeuAspAsnGlyTrpLeuTyr 159
              |||||||||||||||||||||||||||||||      |||||||||||||||||||
Db        307 GCATCTTGGGACCGGATCAGACACTACAGGATACAGCGTCTTGACAATGGCTGGCTGTAC 366

Qy        160 IleSerPro 162
              |||||||||
Db        367 ATCTCACCT 375
```

RESULT 14
BB635615
LOCUS BB635615 660 bp mRNA linear EST 26-OCT-2001
DEFINITION BB635615 RIKEN full-length enriched, 0 day neonate thymus Mus musculus cDNA clone A430076N07 5', mRNA sequence.
ACCESSION BB635615
VERSION BB635615.1 GI:16471660
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 660)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. .

10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .660
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A430076N07"
/tissue_type="thymus"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 0 day neonate thymus"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Alignment Scores:

Pred. No.:	2.51e-45	Length:	660
Score:	560.50	Matches:	115
Percent Similarity:	87.23%	Conservative:	8
Best Local Similarity:	81.56%	Mismatches:	17
Query Match:	41.49%	Indels:	1
DB:	2	Gaps:	1

US-10-043-649-2 (1-261) x BB635615 (1-660)

```
Qy          1 MetGlySerLeuProSerArgArgLysSerLeuProSerProSerLeuSerSerSerVal 20
              |||||||||||||   ||||||   |||:::     |||||||||   |||||||||
Db        241 ATGGGAAGTTTGTCCAGCAGAGGGAAAACC---TCCAGCCCCAGCCCCAGCTCCTCTGGT 297

Qy         21 GlnGlyGlnGlyProValThrMetGluAlaGluArgSerLysAlaThrAlaValAlaLeu 40
                 |||   ||||||:::|||:::   ||||||   |||   ||||||||||||||||
Db        298 CCAGACCAGGAACCCGTGTCCATGCAACCAGAAAGACACAAGGTCACAGCTGTGGCCCTG 357

Qy         41 GlySerPheProAlaGlyGlyProAlaGluLeuSerLeuArgLeuGlyGluProLeuThr 60
              ||||||||||||||||||      |||   ||||||||||||||||||||||||||||||
Db        358 GGCAGTTTCCCAGCAGGTGAACAGGCCAGACTATCTCTGAGACTCGGGGAGCCGCTGACC 417

Qy         61 IleValSerGluAspGlyAspTrpTrpThrValLeuSerGluValSerGlyArgGluTyr 80
              |||:::|||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db        418 ATCATCTCTGAGGATGGAGATTGGTGGACAGTCCAGTCGGAAGTCTCAGGCAGAGAGTAC 477

Qy         81 AsnIleProSerValHisValAlaLysValSerHisGlyTrpLeuTyrGluGlyLeuSer 100
              ::::::|||||||||:::|||||||||||:::|||||||||||||||||||||||||||
Db        478 CACATGCCCAGTGTGTATGTGGCTAAAGTCGCCCACGGGTGGCTGTACGAGGGCCTGAGC 537

Qy        101 ArgGluLysAlaGluGluLeuLeuLeuProGlyAsnProGlyGlyAlaPheLeuIle 120
              ||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db        538 CGGGAGAAAGCCGAGGAACTACTCCTGTTACCTGNGAACCCCGGAGGGGCCTTCCTCATC 597

Qy        121 ArgGluSerGlnThrArgArgGlySerTyrSerLeuSerValArgLeuSerArgProAla 140
              ||||||||||||||||||||||||   |||||||||||||||||||||||||||||||
Db        598 CGGGAGAGCCAGACCAGGAGAGGCTGCTATTCCCTGTCCGTTCGACTCAGCCGCCCTGCA 657

Qy        141 Ser 141
              |||
Db        658 TCT 660
```

RESULT 15
AL844311
LOCUS AL844311 597 bp mRNA linear EST 30-JUL-2002
DEFINITION AL844311 pool_YT_lib_v_SPD Homo sapiens cDNA, mRNA sequence.
ACCESSION AL844311
VERSION AL844311.1 GI:22019093
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 597)
AUTHORS Ashcroft,K., Bethel,G., Bye,J.M., Howell,G.R., Huckle,E.J. and Sheridan,E.
TITLE Homo sapiens EST sequence
JOURNAL Unpublished (2002)
COMMENT Contact: The Sanger Centre
The Sanger Centre
Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK
Email: humquery@sanger.ac.uk
Sanger Centre name : sccd10818.154136A
Homo sapiens EST sequence. This sequence was generated as part of The Wellcome Trust Sanger Institute program to identify and annotate genes in the human genome. Incomplete or unconfirmed genes are experimentally analysed using a variety of cDNA library resources. This sequence was obtained from a PCR product generated from a pool of up to 100,000 cDNA clones derived from pool_YT_lib_v_SPD cDNA library. Further information can be found at http://www.sanger.ac.uk/Teams/Team69/.

FEATURES Location/Qualifiers
source 1. .597
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="20"
/clone_lib="pool_YT_lib_v_SPD"

ORIGIN

Alignment Scores:

Pred. No.:	7.64e-45	Length:	597
Score:	555.00	Matches:	107
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	41.08%	Indels:	0
DB:	1	Gaps:	0

US-10-043-649-2 (1-261) x AL844311 (1-597)

```
Qy        155 AsnGlyTrpLeuTyrIleSerProArgLeuThrPheProSerLeuGlnAlaLeuValAsp 174
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 AATGGCTGGCTGTACATCTCACCACGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGAC 72

Qy        175 HisTyrSerGluLeuAlaAspAspIleCysCysLeuLeuLysGluProCysValLeuGln 194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 CATTACTCTGAGCTGGCGGATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAG 132

Qy        195 ArgAlaGlyProLeuProGlyLysAspIleProLeuProValThrValGlnArgThrPro 214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 AGGGCTGGCCCGCTCCCTGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCA 192
```

```
Qy        215 LeuAsnTrpLysGluLeuAspSerSerLeuLeuPheSerGluAlaAlaThrGlyGluGlu 234
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 CTCAACTGGAAAGAGCTGGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAG 252

Qy        235 SerLeuLeuSerGluGlyLeuArgGluSerLeuSerPheTyrIleSerLeuAsnAspGlu 254
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        253 TCTCTTCTCAGTGAGGGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAG 312

Qy        255 AlaValSerLeuAspAspAla 261
              ||||||||||||||||||||||
Db        313 GCTGTCTCTTTGGATGATGCC 333
```

Search completed: November 17, 2004, 03:11:31
Job time : 3383 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 16, 2004, 20:44:19 ; Search time 3247 Seconds
(without alignments)
8820.951 Million cell updates/sec

Title: US-10-043-649-1
Perfect score: 786
Sequence: 1 atgggaagtctgcccagcag.....tctctttggatgatgcctag 786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	734.6	93.5	1002	5	BQ052308	BQ052308 AGENCOURT
	2	654	83.2	1020	5	BQ054281	BQ054281 AGENCOURT
	3	645	82.1	1069	5	BQ052468	BQ052468 AGENCOURT
	4	537.8	68.4	2637	3	AK088672	AK088672 Mus muscu
	5	536.2	68.2	2974	3	AK030877	AK030877 Mus muscu
	6	529	67.3	1032	1	AL541041	AL541041 AL541041
	7	478	60.8	926	3	AK020837	AK020837 Mus muscu
	8	453.8	57.7	660	6	BY742155	BY742155 BY742155
	9	405	51.5	566	4	BG284179	BG284179 602408226
	10	386.2	49.1	986	5	BQ054265	BQ054265 AGENCOURT
	11	331.2	42.1	878	5	BQ053486	BQ053486 AGENCOURT
	12	326.8	41.6	778	4	BG178487	BG178487 602328305
	13	322.4	41.0	597	1	AL844311	AL844311 AL844311
	14	321	40.8	960	5	BU944126	BU944126 AGENCOURT
	15	317.4	40.4	614	1	AL844307	AL844307 AL844307
	16	314.6	40.0	606	1	AL844309	AL844309 AL844309
	17	296.2	37.7	660	2	BB635615	BB635615 BB635615
	18	289.8	36.9	762	5	BP162888	BP162888 BP162888
	19	283.6	36.1	377	1	AA959151	AA959151 vz51g06.r
	20	283.2	36.0	701	7	CK833360	CK833360 4057315 B
	21	283	36.0	781	7	CK596391	CK596391 AGENCOURT
c	22	277.6	35.3	569	6	CB426333	CB426333 601508 MA
	23	275.6	35.1	627	2	BB619854	BB619854 BB619854
	24	261.4	33.3	794	4	BG677567	BG677567 602624118
	25	249.4	31.7	656	7	CN792083	CN792083 4126904 B
c	26	221.6	28.2	603	5	BQ553005	BQ553005 H4019E02-
	27	220	28.0	670	1	AI510095	AI510095 mj43c04.y
	28	215.8	27.5	640	5	BQ553006	BQ553006 H4019E02-
	29	196.2	25.0	575	5	BP156169	BP156169 BP156169
	30	175.6	22.3	701	5	BX849096	BX849096 BX849096
	31	170.8	21.7	569	6	CB514688	CB514688 ssalrgb52
c	32	166.8	21.2	642	1	AL844308	AL844308 AL844308
	33	163.4	20.8	831	9	AY416277	AY416277 Homo sapi
	34	162.2	20.6	871	5	BQ436143	BQ436143 AGENCOURT
c	35	161.8	20.6	619	1	AL844312	AL844312 AL844312
	36	160.6	20.4	570	1	AI471720	AI471720 ta16a01.x
	37	160.6	20.4	1762	3	CR617843	CR617843 full-leng
	38	160.2	20.4	882	5	BX415149	BX415149 BX415149
	39	160.2	20.4	1015	5	BX436423	BX436423 BX436423
	40	160.2	20.4	1037	1	AL539427	AL539427 AL539427
	41	160.2	20.4	1054	1	AL549826	AL549826 AL549826
	42	159.8	20.3	966	1	AL551370	AL551370 AL551370
	43	159.2	20.3	1063	5	BQ072745	BQ072745 AGENCOURT
	44	159	20.2	972	5	BQ707614	BQ707614 AGENCOURT
	45	158	20.1	562	5	BX383606	BX383606 BX383606

ALIGNMENTS

RESULT 1
BQ052308
LOCUS BQ052308 1002 bp mRNA linear EST 29-MAR-2002
DEFINITION AGENCOURT_6868571 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:5933542 5', mRNA sequence.
ACCESSION BQ052308
VERSION BQ052308.1 GI:19811648
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1002)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2118 row: d column: 23
High quality sequence stop: 670.
FEATURES Location/Qualifiers
source 1. .1002
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5933542"
/tissue_type="natural killer cells, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_106"
/note="Organ: blood; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
ORIGIN

Query Match 93.5%; Score 734.6; DB 5; Length 1002;

```
  Best Local Similarity   98.3%;  Pred. No. 4.8e-183;
  Matches  774;  Conservative    0;  Mismatches    9;  Indels    4;  Gaps    3;

Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         96 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 155

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        156 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 215

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        216 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 275

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        276 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 335

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        336 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 395

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 455

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        456 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 515

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        516 TCCTGGGACCGGATCAGACACTACAGGATCTACTGCCTTGACAATGGCTGGCTGTACATC 575

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        576 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 635

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        636 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 695

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        696 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 755

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAGCTGCCACA-GGGGAGGAGTCTCTTCTCAGTGAGGG 719
              |||||||||||||||||||||||| |||||||| ||||||||||||||||||||||||||
Db        756 GACAGCTCCCTCCTGTTTTCTGAAACTGCCACAGGGGGAGGAGTCTCTTCTCAGTGAGGG 815

Qy        720 --TCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGA-GGCTGTCTCTTTGGA 776
                  | ||||||||||||||||||||||||||||   ||||||| ||||||||||||||||
Db        816 GTCTCCGGGGAGTCCCTCAGCTTCTACATCAGCCCTGATGACGAGGGCTGTCTCTTTGGA 875

Qy        777 TGATGCC 783
              |||||||
Db        876 TGATGCC 882


RESULT 2
BQ054281
LOCUS       BQ054281               1020 bp    mRNA    linear   EST 29-MAR-2002
DEFINITION  AGENCOURT_6830234 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:5936362
            5', mRNA sequence.
ACCESSION   BQ054281
VERSION     BQ054281.1  GI:19813621
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1020)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2125  row: j  column: 11
            High quality sequence stop: 556.
FEATURES             Location/Qualifiers
     source          1. .1020
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5936362"
                     /tissue_type="natural killer cells, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_106"
                     /note="Organ: blood; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN

  Query Match             83.2%;  Score 654;  DB 5;  Length 1020;
  Best Local Similarity   95.4%;  Pred. No. 9.9e-162;
  Matches  748;  Conservative    0;  Mismatches   27;  Indels    9;  Gaps    7;

Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        212 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 271

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        272 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 331

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        332 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 391

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        392 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 451

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        452 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 511

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        512 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGNGAACCCTGGAGGGGCCTTCCTCATC 571

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        572 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 631

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        632 TCCTGNGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 691

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGG-CCCTGGTGGACCATTACTCTGAGCTGGC 539
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db        692 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCCTGGTGGACCATTACTCTGAGCTGGC 751
```

```
Qy        540 GGATGACATCTGCTGCCTA-CTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCC 598
              |||||||||||||||||||| |||| ||| |||||||||||||| ||||||| ||||||||
Db        752 GGATGACATCTGCTGCCTANCTCAGGGACCCCTGTGTCCTGCAAAGGGCTGCCCCGCTCC 811

Qy        599 CTGGCAAGGATATA-CCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAG 657
              ||||||| |||||| |||||||||||||||||||| |||||||      ||||| |||||
Db        812 CTGGCAAAGATATACCCCCTACCTGTGACTGTGCAAAGGACACCCACTCACTGGGAAGAG 871

Qy        658 CTGGACAGCTCCCTCCTG--TTTTCTGAAGCTGCCACA-GGGGAGGAGTCTCTTCTCAGT 714
              |||||||||||||||||   |||  |||||||||| || |||||||||||||||||| |
Db        872 CTGGACAGCTCCCTCCTGGTTTTCTTGAAGCTGCCCCAGGGGGAGGAGTCTCTTCTCATT 931

Qy        715 GAGGG--TCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTG-AATGACGAGGCTGTCTCT 771
              |||||    |||||||||||||||||||||||||| |||| |||||| |||||| |
Db        932 GAGGGGTCTTCCGGGAGTCCCTCAGCTTCTACATCAACCTGTAATGACCAGGCTGGCCTC 991

Qy        772 TTGG 775
              || |
Db        992 TTTG 995
```

RESULT 3
BQ052468
LOCUS BQ052468 1069 bp mRNA linear EST 29-MAR-2002
DEFINITION AGENCOURT_6868422 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:5933772 5', mRNA sequence.
ACCESSION BQ052468
VERSION BQ052468.1 GI:19811808
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1069)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2118 row: n column: 13
High quality sequence stop: 681.
FEATURES Location/Qualifiers
source 1. .1069
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5933772"
/tissue_type="natural killer cells, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_106"
/note="Organ: blood; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
ORIGIN

Query Match 82.1%; Score 645; DB 5; Length 1069;
Best Local Similarity 93.4%; Pred. No. 2.4e-159;
Matches 737; Conservative 0; Mismatches 0; Indels 52; Gaps 4;

```
Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         89 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 148

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        149 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 208

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        209 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 268

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        269 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 328

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        329 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 388

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        389 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 448

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        449 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 508

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        509 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 568

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCT-GGTGGACCATTACTCTGAGCTGGC 539
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        569 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGGTGGACCATTACTCTGAG----- 623

Qy        540 GGATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCC 599
                                                  ||||||||||||||||
Db        624 ----------------------------------------GGGCTGGCCCGCTCCC 639

Qy        600 TGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCT 659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 TGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCT 699

Qy        660 GGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACA-GGGGAGGAGTCTCTTCTCAGTGAGG 718
              |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db        700 GGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGGAGGAGTCTCTTCTCAGTGAGG 759

Qy        719 GTCTCCGGG-AGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGAT 777
              ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 GTCTCCGGGAAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGAT 819

Qy        778 GATGCCTAG 786
              |||||||||
Db        820 GATGCCTAG 828
```

RESULT 4
AK088672
LOCUS AK088672 2637 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN full-length enriched library, clone:E430023D24 product:MODULATOR OF ANTIGEN RECEPTOR SIGNALING MARS, full insert sequence.
ACCESSION AK088672
VERSION AK088672.1 GI:26353729
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253

PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2637)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/.
FEATURES Location/Qualifiers
source 1. .2637
/organism="Mus musculus"
/mol_type="mRNA"
/strain="NOD"
/db_xref="FANTOM_DB:E430023D24"
/db_xref="taxon:10090"
/clone="E430023D24"
/cell_type="thymic cells"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="2 days neonate"
CDS 358. .1137
/note="unnamed protein product; MODULATOR OF ANTIGEN RECEPTOR SIGNALING MARS (SPTR|AAL38196, evidence: FASTY, 100%ID, 100%length, match=777)
putative"
/codon_start=1
/protein_id="BAC40495.1"
/db_xref="GI:26353730"
/translation="MGSLSSRGKTSSPSPSSSGPDQEPVSMQPERHKVTAVALGSFPA
GEQARLSLRLGEPLTIISEDGDWWTVQSEVSGREYHMPSVYVAKVAHGWLYEGLSREK
AEELLLLPGNPGGAFLIRESQTRRGCYSLSVRLSRPASWDRIRHYRIQRLDNGWLYIS
PRLTFPSLHALVEHYSELADGICCPLREPCVLQKLGPLPGKDTPPPVTVPTSSLNWKK
LDRSLLFLEAPASGEASLLSEGLRESLSSYISLAEDPLDDA"
ORIGIN

Query Match 68.4%; Score 537.8; DB 3; Length 2637;
Best Local Similarity 81.9%; Pred. No. 6.5e-131;
Matches 646; Conservative 0; Mismatches 137; Indels 6; Gaps 2;

```
Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              |||||||| || |||||||| | ||| |     | ||||| |||   || |||||||
Db        358 ATGGGAAGTTTGTCCAGCAGAGGGAAAAC---CTCCAGCCCCAGCCCCAGCTCCTCTGGT 414

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              | || ||||| ||| ||| ||||| || |||| ||| |||||  |||||||| |||||||||
Db        415 CCAGACCAGGAACCCGTGTCCATGCAACCAGAAAGACACAAGGTCACAGCTGTGGCCCTG 474

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              |||||||||| |||||||  | ||||   || || ||||||||||||||||  |||||
Db        475 GGCAGTTTCCCAGCAGGTGAACAGGCCAGACTATCTCTGAGACTCGGGGAGCCGCTGACC 534

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||| |||||||||||||||| |||||||| || | ||| ||||||||||||||||||||
Db        535 ATCATCTCTGAGGATGGAGATTGGTGGACAGTCCAGTCGGAAGTCTCAGGCAGAGAGTAC 594

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              |||| ||||| ||  | ||||| |||||| |||| |||||||||||| ||||||||||||
Db        595 CACATGCCCAGTGTGTATGTGGCTAAAGTCGCCCACGGGTGGCTGTACGAGGGCCTGAGC 654

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              |||||||||| |||||||| ||  |||||||||||||||| ||||||||||||||||||
Db        655 CGGGAGAAAGCCGAGGAACTACTCCTGTTACCTGGGAACCCCGGAGGGGCCTTCCTCATC 714

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              |||||||||||||||||||||||||  || || ||||| ||||| ||||||||||||||
Db        715 CGGGAGAGCCAGACCAGGAGAGGCTGCTATTCCCTGTCCGTCCGACTCAGCCGCCCTGCA 774

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              || |||||||||||||||||||||||| ||  | ||||||||||||||||||||||||
Db        775 TCTTGGGACCGGATCAGACACTACAGGATACAGCGTCTTGACAATGGCTGGCTGTACATC 834

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              ||||| |||||||||||||||||||||| ||| ||||||| |||||||||||||| ||
Db        835 TCACCTCGCCTCACCTTCCCCTCACTCCACGCCTTGGTGGAGCATTACTCTGAGCTAGCA 894

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              |||| |||||||||| |  |||| |||||| |||||||||||| |  ||| || || |||
Db        895 GATGGCATCTGCTGTCCCCTCAGGGAGCCGTGTGTCCTGCAGAAGCTTGGGCCACTACCT 954

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||| |||| ||| | |||||||||||||| |   || |||| || |||||| |||||
Db        955 GGCAAAGATACACCTCCACCTGTGACTGTGCCAACATCATCACTAAATTGGAAAAAGCTG 1014

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAG---CTGCCACAGGGGAGGAGTCTCTTCTCAGTGAG 717
              ||| ||  |||||||||  ||||   |||| | ||||||  ||||| |||||||||
Db       1015 GACCGCAGCCTCCTGTTTCTGGAAGCACCTGCGAGTGGGGAGGCATCTCTGCTCAGTGAG 1074

Qy        718 GGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGAT 777
```

```
Db       1075 GGGCTCCGAGAGTCCCTCAGTTCCTACATCAGCCTGGCTGAGGACCCCTTGGATGATGCT 1134

Qy        778 GATGCCTAG 786
              |  ||| |
Db       1135 TAGCCCTGG 1143
```

RESULT 5
AK030877
LOCUS AK030877 2974 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus adult male thymus cDNA, RIKEN full-length enriched library, clone:5830437K10 product:MODULATOR OF ANTIGEN RECEPTOR SIGNALING MARS, full insert sequence.
ACCESSION AK030877
VERSION AK030877.1 GI:26326848
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2974)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/.
FEATURES Location/Qualifiers
source 1. .2974
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:5830437K10"
/db_xref="taxon:10090"
/clone="5830437K10"
/sex="male"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
CDS 232. .1011
/note="unnamed protein product; MODULATOR OF ANTIGEN RECEPTOR SIGNALING MARS (SPTR|AAL38196, evidence: FASTY, 100%ID, 100%length, match=777)
putative"
/codon_start=1
/protein_id="BAC27168.1"
/db_xref="GI:26326849"
/translation="MGSLSSRGKTSSPSPSSSGPDQEPVSMQPERHKVTAVALGSFPA
GEQARLSLRLGEPLTIISEDGDWWTVQSEVSGREYHMPSVYVAKVAHGWLYEGLSREK
AEELLLLPGNPGGAFLIRESQTRRGYYSLSVRLSRPASWDRIRHYRIQRLDNGWLYIT
PRLTFPSLHALVEHYSELADGICCPLREPCVLQKLGPLPGKDTPPPVTVPTSSLNWKK
LDRSLLFLEAPASGEASLLSEGLRESLSSYISLAEDPLDDA"
ORIGIN

Query Match 68.2%; Score 536.2; DB 3; Length 2974;
Best Local Similarity 81.7%; Pred. No. 1.8e-130;
Matches 645; Conservative 0; Mismatches 138; Indels 6; Gaps 2;

```
Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              |||||||||  ||  |||||||||  |  |||  |      |  |||||  |||   ||  |||||||
Db        232 ATGGGAAGTTTGTCCAGCAGAGGGAAAAC---CTCCAGCCCCAGCCCCAGCTCCTCTGGT 288

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              |  ||  |||||  |||  |||  |||||  ||  ||||  |||   |||||  ||||||  |||||||||
Db        289 CCAGACCAGGAACCCGTGTCCATGCAACCAGAAAGACACAAGGTCACAGCTGTGGCCCTG 348

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||  |||||||   |  ||||   ||  ||  ||||||||||||||||||   |||||
Db        349 GGCAGTTTCCCAGCAGGTGAACAGGCCAGACTATCTCTGAGACTCGGGGAGCCGCTGACC 408

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              |||  ||||||||||||||||  ||||||||  ||  |  |||  ||||||||||||||||||||
Db        409 ATCATCTCTGAGGATGGAGATTGGTGGACAGTCCAGTCGGAAGTCTCAGGCAGAGAGTAC 468

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||  |||||  ||  |  |||||  ||||||  ||||  ||||||||||  ||||||||||||
Db        469 CACATGCCCAGTGTGTATGTGGCTAAAGTCGCCCACGGGTGGCTGTACGAGGGCCTGAGC 528

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||  ||||||||  ||  ||||||||||||||||  ||||||||||||||||||
Db        529 CGGGAGAAAGCCGAGGAACTACTCCTGTTACCTGGGAACCCCGGAGGGGCCTTCCTCATC 588

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              ||||||||||||||||||||||||  ||  ||  |||||  |||||  ||||||||||||||||
Db        589 CGGGAGAGCCAGACCAGGAGAGGCTACTATTCCCTGTCCGTCCGACTCAGCCGCCCTGCA 648
```

```
Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              || |||||||||||||||||||||||||||| ||  | |||||||||||||||||||||||
Db        649 TCTTGGGACCGGATCAGACACTACAGGATACAGCGTCTTGACAATGGCTGGCTGTACATC 708

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
               |||| |||||||||||||||||||||||| ||| ||||||| |||||||||||||| ||
Db        709 ACACCTCGCCTCACCTTCCCCTCACTCCACGCCTTGGTGGAGCATTACTCTGAGCTAGCA 768

Qy        541 GATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCT 600
              |||| ||||||||| |  |||| |||||| |||||||||||| |  ||| || || |||
Db        769 GATGGCATCTGCTGTCCCCTCAGGGAGCCGTGTGTCCTGCAGAAGCTTGGGCCACTACCT 828

Qy        601 GGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTG 660
              ||||| |||| ||| | |||||||||||||  |   || |||| || ||||||| |||||
Db        829 GGCAAAGATACACCTCCACCTGTGACTGTGCCAACATCATCACTAAATTGGAAAAAGCTG 888

Qy        661 GACAGCTCCCTCCTGTTTTCTGAAG---CTGCCACAGGGGAGGAGTCTCTTCTCAGTGAG 717
              ||| ||  |||||||||||   ||||  |||| |  ||||||| |||||  |||||||||
Db        889 GACCGCAGCCTCCTGTTTCTGGAAGCACCTGCGAGTGGGGAGGCATCTCTGCTCAGTGAG 948

Qy        718 GGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGAT 777
              || ||||| |||||||||| | |||||||||||||  ||| ||  |  |   |  | ||
Db        949 GGGCTCCGAGAGTCCCTCAGTTCCTACATCAGCCTGGCTGAGGACCCCTTGGATGATGCT 1008

Qy        778 GATGCCTAG 786
                |  ||| |
Db       1009 TAGCCCTGG 1017
```

RESULT 6
AL541041

```
LOCUS       AL541041                1032 bp    mRNA    linear   EST 24-MAR-2004
DEFINITION  AL541041 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE005YK23
            5-PRIME, mRNA sequence.
ACCESSION   AL541041
VERSION     AL541041.3  GI:45716635
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1032)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 15, 2001 this sequence version replaced gi:30544829.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 9825.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0DE005AF12QP1&c=9825.r.
FEATURES             Location/Qualifiers
     source          1. .1032
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE005YK23"
                     /tissue_type="PLACENTA"
                     /clone_lib="Homo sapiens PLACENTA"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN
```

Query Match 67.3%; Score 529; DB 1; Length 1032;
Best Local Similarity 96.5%; Pred. No. 1.1e-128;
Matches 526; Conservative 12; Mismatches 7; Indels 0; Gaps 0;

```
Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        395 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 454

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        455 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 514

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        515 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 574

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        575 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 634

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        635 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 694

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        695 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCATC 754

Qy        361 CGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 420
              |||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db        755 CGGGAGAGCCAGAMCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCA 814

Qy        421 TCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 480
              |||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db        815 TCCTGGGACCGGATCAGACACTAMAGGATCCACTGCCTTGACAATGGCTGGCTGTACATC 874

Qy        481 TCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCG 540
              |||| ||||||:| |:|:|||:|||||||||||||||||||:||||||:|:|:  || |
Db        875 TCACGGCGCCTYAACYTYCCCYCACTCCAGGCCCTGGTGGACMATTACTYTRASTGGGGG 934

Qy        541 GATGA 545
                : : |
Db        935 RTKAA 939
```

RESULT 7
AK020837

```
LOCUS       AK020837                926 bp    mRNA    linear   HTC 03-APR-2004
DEFINITION  Mus musculus adult retina cDNA, RIKEN full-length enriched library,
            clone:A930009E21 product:MODULATOR OF ANTIGEN RECEPTOR SIGNALING
            MARS, full insert sequence.
ACCESSION   AK020837
VERSION     AK020837.1  GI:12861542
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
```

REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 926)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through two rounds of normalization to Rot = 20.0 and subtraction to Rot = 458.8. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B. Retina RNA was provided by Stefano Gustincich (Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA 02115, USA) whose assistance is gratefully acknowledged.
Retina RNA was provided by Stefano Gustincich (Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA 02115, USA) whose assistance is gratefully acknowledged.

FEATURES Location/Qualifiers
source 1. .926
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:A930009E21"
/db_xref="taxon:10090"
/clone="A930009E21"
/tissue_type="retina"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
CDS 128. .667
/note="unnamed protein product; MODULATOR OF ANTIGEN RECEPTOR SIGNALING MARS (SPTR|AAL38196, evidence: FASTY, 100%ID, 100%length, match=777)
putative"
/codon_start=1
/protein_id="BAB32223.1"
/db_xref="GI:12861543"
/translation="MPSVYVAKVAHGWLYEGLSREKAEELLLLPGNPGGAFLIRESQT RRGCYSLSVRLSRPASWDRIRHYRIQRLDNGWLYISPRLTFPSLHALVEHYSELADGI CCPLREPCVLQKLGPLPGKDTPPPVTVPTSSLNWKKLDRSLLFLEAPASGEASLLSEG LRESLSSYISLAEDPLDDA"
polyA_signal 907. .912
/note="putative"
polyA_site 926
/note="putative"

ORIGIN

Query Match 60.8%; Score 478; DB 3; Length 926;
Best Local Similarity 83.1%; Pred. No. 3.3e-115;
Matches 557; Conservative 0; Mismatches 110; Indels 3; Gaps 1;

```
Qy        120 GGGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGAC 179
              ||||||||||| |||||||  | ||||   || || ||||||||||||||||||  ||||
Db          4 GGGCAGTTTCCCAGCAGGTGAACAGGCCAGACTATCTCTGAGACTCGGGGAGCCGCTGAC 63

Qy        180 CATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTA 239
              |||| |||||||||||||||| |||||||| || | ||| ||||||||||||||||||||
Db         64 CATCATCTCTGAGGATGGAGATTGGTGGACAGTCCAGTCGGAAGTCTCAGGCAGAGAGTA 123

Qy        240 TAACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAG 299
              |||| ||||| ||  | ||||| |||||| |||| |||||||||| ||||||||||
Db        124 CCACATGCCCAGTGTGTATGTGGCTAAAGTCGCCCACGGGTGGCTGTACGAGGGCCTGAG 183

Qy        300 CAGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCAT 359
              | |||||||||| |||||||| ||  |||||||||||||||| ||||||||||||||||
Db        184 CCGGGAGAAAGCCGAGGAACTACTCCTGTTACCTGGGAACCCCGGAGGGGCCTTCCTCAT 243

Qy        360 CCGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGC 419
              |||||||||||||||||||||||||||||||  || || ||||| ||||| ||||||||||||||
Db        244 CCGGGAGAGCCAGACCAGGAGAGGCTGCTATTCCCTGTCCGTCCGACTCAGCCGCCCTGC 303

Qy        420 ATCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACAT 479
              ||| |||||||||||||||||||||||||||||| ||  | ||||||||||||||||||||||
Db        304 ATCTTGGGACCGGATCAGACACTACAGGATACAGCGTCTTGACAATGGCTGGCTGTACAT 363

Qy        480 CTCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGC 539
              |||||| |||||||||||||||||||||||| ||| ||||||| |||||||||||||| ||
Db        364 CTCACCTCGCCTCACCTTCCCCTCACTCCACGCCTTGGTGGAGCATTACTCTGAGCTAGC 423

Qy        540 GGATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCC 599
              |||| |||||||| |  |||| |||||| |||||||||||| |  ||| || || ||
Db        424 AGATGGCATCTGCTGTCCCCTCAGGGAGCCGTGTGTCCTGCAGAAGCTTGGGCCACTACC 483

Qy        600 TGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCT 659
              |||||| |||| ||| | |||||||||||||| |  || |||| || |||||| ||||
Db        484 TGGCAAAGATACACCTCCACCTGTGACTGTGCCAACATCATCACTAAATTGGAAAAAGCT 543

Qy        660 GGACAGCTCCCTCCTGTTTTCTGAAG---CTGCCACAGGGGAGGAGTCTCTTCTCAGTGA 716
              |||| ||  ||||||||||  ||||  |||| | ||||||| |||||||||||||||
Db        544 GGACCGCAGCCTCCTGTTTCTGGAAGCACCTGCGAGTGGGGAGGCATCTCTGCTCAGTGA 603

Qy        717 GGGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGGA 776
              ||| ||||| |||||||||| | |||||||||||||  ||| ||  |  |  |  |
Db        604 GGGGCTCCGAGAGTCCCTCAGTTCCTACATCAGCCTGGCTGAGGACCCCTTGGATGATGC 663

Qy        777 TGATGCCTAG 786
```

```
                  | |  ||| |
Db        664 TTAGCCCTGG 673


RESULT 8
BY742155
LOCUS       BY742155                660 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION  BY742155 RIKEN full-length enriched, adult retina Mus musculus cDNA
            clone A930009E21 5', mRNA sequence.
ACCESSION   BY742155
VERSION     BY742155.1  GI:27167493
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 660)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P.,
            Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y.,
            Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M.,
            Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N.,
            Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M.,
            Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
            Direct Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Retina RNA was provided by Dr. Stefano Gustincich (Department of
            Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA
            02115, USA) whose assistance is gratefully acknowledged.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .660
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="A930009E21"
                     /tissue_type="retina"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult retina"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3']. cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5'GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. ~Retina RNA was provided by Stefano Gustincich,
                     Department of Neurobiology, Harvard Medical School, 220
                     Longwood Ave., Boston, MA02115, USA, whose assistance we
                     gratefully acknowledge."
ORIGIN

  Query Match             57.7%;  Score 453.8;  DB 6;  Length 660;
  Best Local Similarity   83.6%;  Pred. No. 7.7e-109;
  Matches  539;  Conservative    0;  Mismatches  102;  Indels    4;  Gaps    2;

Qy        120 GGGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGAC 179
              ||||||||||| |||||||  | ||||   || || ||||||||||||||||||  ||||
Db          4 GGGCAGTTTCCCAGCAGGTGAACAGGCCAGACTATCTCTGAGACTCGGGGAGCCGCTGAC 63

Qy        180 CATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTA 239
              |||| |||||||||||||||| |||||||| || | ||| ||||||||||||||||||||
Db         64 CATCATCTCTGAGGATGGAGATTGGTGGACAGTCCAGTCGGAAGTCTCAGGCAGAGAGTA 123

Qy        240 TAACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAG 299
              |||| ||||| ||  | ||||| |||||| |||| |||||||||| ||||||||||
Db        124 CCACATGCCCAGTGTGTATGTGGCTAAAGTCGCCCACGGGTGGCTGTACGAGGGCCTGAG 183

Qy        300 CAGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCAT 359
              | |||||||||| |||||||| ||  |||||||||||||||| ||||||||||||||||
Db        184 CCGGGAGAAAGCCGAGGAACTACTCCTGTTACCTGGGAACCCCGGAGGGGCCTTCCTCAT 243

Qy        360 CCGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGC 419
              |||||||||||||||||||||||||  || || ||||| ||||| ||||||||||||||
Db        244 CCGGGAGAGCCAGACCAGGAGAGGCTGCTATTCCCTGTCCGTCCGACTCAGCCGCCCTGC 303

Qy        420 ATCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGTACAT 479
              ||| ||||||||||||||||||||||||||  || | ||||||||||||||||||||||
Db        304 ATCTTGGGACCGGATCAGACACTACAGGATACAGCGTCTTGACAATGGCTGGCTGTACAT 363
```

```
Qy        480 CTCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGC 539
              |||||| |||||||||||||||||||||||| ||| ||||||| ||||||||||||||| ||
Db        364 CTCACCTCGCCTCACCTTCCCCTCACTCCACGCCTTGGTGGAGCATTACTCTGAGCTAGC 423

Qy        540 GGATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCC 599
              ||||  |||||||||  |   |||| ||||||  |||||||||||||  |  ||| || || ||
Db        424 AGATGGCATCTGCTGTCCCCTCAGGGAGCCGTGTGTCCTGCAGAAGCTTGGGCCACTACC 483

Qy        600 TGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCT 659
              |||||| |||| ||| | ||||||||||||||   |    || |||| || |||||| ||||
Db        484 TGGCAAAGATACACCTCCACCTGTGACTGTGCCAACATCATCACTAAATTGGAAAAAGCT 543

Qy        660 GGACAGCTCCCTCCTGTTTTCTGAAG---CTGCCACAGGGGAGGAGTCTCTTCTCAGTGA 716
              |||| |   |||||||||   ||||   |||| |  |||||||  ||||| |||| |||
Db        544 GGACCGTAGCCTCCTGTTTCTGGAAGCACCTGCAAGTGGGGAGGCATCTCTGCTCATTGA 603

Qy        717 GGGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGA 761
              ||| |   |  |||| |||||  |||||||| ||| || |  ||| ||
Db        604 GGGGCCTCCAGAGT-CCTCACCTTCTACCTCATCCCGGTTGATGA 647
```

RESULT 9
BG284179
LOCUS BG284179 566 bp mRNA linear EST 21-FEB-2001
DEFINITION 602408226F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4520382 5', mRNA sequence.
ACCESSION BG284179
VERSION BG284179.1 GI:13034866
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 566)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10418 row: c column: 07
High quality sequence start: 2
High quality sequence stop: 566.

```
FEATURES             Location/Qualifiers
     source          1. .566
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4520382"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_91"
                     /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.4 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
ORIGIN
```

Query Match 51.5%; Score 405; DB 4; Length 566;
Best Local Similarity 100.0%; Pred. No. 6e-96;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        382 GGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCATCCTGGGACCGGATCAGACAC 441
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 GGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCCCTGCATCCTGGGACCGGATCAGACAC 72
```

```
Qy        442 TACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATCTCACCGCGCCTCACCTTCCCC 501
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 TACAGGATCCACTGCCTTGACAATGGCTGGCTGTACATCTCACCGCGCCTCACCTTCCCC 132

Qy        502 TCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCGGATGACATCTGCTGCCTACTC 561
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 TCACTCCAGGCCCTGGTGGACCATTACTCTGAGCTGGCGGATGACATCTGCTGCCTACTC 192

Qy        562 AAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCTGGCAAGGATATACCCCTACCT 621
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 AAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGCTCCCTGGCAAGGATATACCCCTACCT 252

Qy        622 GTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTGGACAGCTCCCTCCTGTTTTCT 681
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        253 GTGACTGTGCAGAGGACACCACTCAACTGGAAAGAGCTGGACAGCTCCCTCCTGTTTTCT 312

Qy        682 GAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGTCTCCGGGAGTCCCTCAGCTTC 741
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        313 GAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTGAGGGTCTCCGGGAGTCCCTCAGCTTC 372

Qy        742 TACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGATGCCTAG 786
              |||||||||||||||||||||||||||||||||||||||||||||
Db        373 TACATCAGCCTGAATGACGAGGCTGTCTCTTTGGATGATGCCTAG 417
```

RESULT 10
BQ054265
LOCUS BQ054265 986 bp mRNA linear EST 29-MAR-2002
DEFINITION AGENCOURT_6830248 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:5936339 5', mRNA sequence.
ACCESSION BQ054265
VERSION BQ054265.1 GI:19813605
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 986)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2125 row: i column: 12
High quality sequence stop: 515.

```
FEATURES             Location/Qualifiers
     source          1. .986
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5936339"
                     /tissue_type="natural killer cells, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_106"
                     /note="Organ: blood; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN
```

Query Match 49.1%; Score 386.2; DB 5; Length 986;
Best Local Similarity 98.0%; Pred. No. 6.3e-91;

```
  Matches  444;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    5;

Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 338

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 398

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 458

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        459 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 518

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        519 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 578

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGG-CCTTCCTCAT 359
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db        579 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCCTTCCTCAT 638

Qy        360 CCGGGAG-AGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGT-CCGCCTCAGCCG-CCC 416
              ||||||| |||||||||||||||||||||||||||||||||| |||||||||||| |||
Db        639 CCGGGAGAAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCCGCCTCAGCCGCCCC 698

Qy        417 TGCATCCTGGG--ACCGGATCAGACACTACAGG 447
              |||||||||||  | |||||||| |||||||
Db        699 TGCATCCTGGGGACCGGGATCAGACCCTACAGG 731
```

RESULT 11
BQ053486

```
LOCUS       BQ053486                878 bp    mRNA    linear   EST 29-MAR-2002
DEFINITION  AGENCOURT_6822017 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:5935253
            5', mRNA sequence.
ACCESSION   BQ053486
VERSION     BQ053486.1  GI:19812826
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 878)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2122  row: l  column: 06
            High quality sequence stop: 394.
FEATURES             Location/Qualifiers
     source          1. .878
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5935253"
                     /tissue_type="natural killer cells, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_106"
                     /note="Organ: blood; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN
```

```
  Query Match           42.1%;  Score 331.2;  DB 5;  Length 878;
  Best Local Similarity  93.1%;  Pred. No. 2.1e-76;
  Matches  390;  Conservative    0;  Mismatches   24;  Indels    5;  Gaps    4;

Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        273 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 332

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        333 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 392

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        393 CGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 452

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              |||||||||||||||||||||||||||||||||||||||||||| ||||| ||||||||||
Db        453 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAATCTCACGCAGAGAGTAT 512

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              |||||||||| |||||||||||||||||||||||| ||||||||||||||||||||||
Db        513 AACATCCCCATCGTCCACGTGGCCAAAGTCTCCCATGCGTGGCTGTATGAGGGCTGANCA 572

Qy        301 AGGGAGAAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCCTCAT- 359
              |  || |||  |  | | |||||||||||||||||| |||||||||||||| ||| |
Db        573 A--GAAAAAACCAAAGGACTGCTGTTGTTACCTGGGAAACCTGGAGGGGCCTTTCTCCTC 630

Qy        360 CCGGGAGAGCCAGACCAGGAGAGGCT-CTTACTCTCTGT-CAGTCCGCCTCAGCCGCCC 416
              |||||||| |||||||||||||||||| || |||||||||  |||||||| ||||||||
Db        631 CCGGGAGAACCAGACCAGGAGAGGCTCCTAACTCTCTGTGTAGTCCGCCTTAGCCGCCC 689
```

RESULT 12
BG178487

```
LOCUS       BG178487                778 bp    mRNA    linear   EST 06-FEB-2001
DEFINITION  602328305F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4429896 5',
            mRNA sequence.
ACCESSION   BG178487
VERSION     BG178487.1  GI:12685190
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 778)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10182  row: i  column: 01
            High quality sequence stop: 657.
FEATURES             Location/Qualifiers
     source          1. .778
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
```

/clone="IMAGE:4429896"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_91"
/note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.4 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

ORIGIN

Query Match 41.6%; Score 326.8; DB 4; Length 778;
Best Local Similarity 94.6%; Pred. No. 3e-75;
Matches 371; Conservative 0; Mismatches 17; Indels 4; Gaps 3;

```
Qy          1 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 ATGGGAAGTCTGCCCAGCAGAAGAAAATCTCTGCCAAGCCCAAGCTTGAGTTCCTCTGTC 421

Qy         61 CAAGGCCAGGGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 CAAGGCCA-GGACCTGTGACCATGGAAGCAGAGAGAAGCAAGGCCACAGCCGTGGCCCTG 480

Qy        121 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGCAGTTTCCCGGCAGGTGGCCCGGCCGAGCTGTCGCTGAGACTCGGGGAGCCATTGACC 540

Qy        181 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATCGTCTCTGAGGATGGAGACTGGTGGACGGTGCTGTCTGAAGTCTCAGGCAGAGAGTAT 600

Qy        241 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATGAGGGCCTGAGC 300
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        601 AACATCCCCAGCGTCCACGTGGCCAAAGTCTCACATGGGTGGCTGTATGAGGGCCTGAGC 660

Qy        301 AGGGAGA-AAGCAGAGGAACTGCTGTTGTTACCTGGGAACCCTGGAGGGGCCTTCC--TC 357
              ||||||| |||||||||||||||||||||||||||||||  ||||||| | |||
Db        661 AGGGAGACAAGCAGAGGAACTGCTGTTGTTACCTGGGAACCTGGGAGGGGGCCTCCCTCA 720

Qy        358 ATCCGGGAGAGCCAGACCAGGAGAGGCTCTTA 389
                  |||| |||  |||||||||   | || ||
Db        721 TCCCGGAAGAAGCAGACCAGGAAGAGGTCCTA 752
```

RESULT 13
AL844311
LOCUS AL844311 597 bp mRNA linear EST 30-JUL-2002
DEFINITION AL844311 pool_YT_lib_v_SPD Homo sapiens cDNA, mRNA sequence.
ACCESSION AL844311
VERSION AL844311.1 GI:22019093
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 597)
AUTHORS Ashcroft,K., Bethel,G., Bye,J.M., Howell,G.R., Huckle,E.J. and
Sheridan,E.
TITLE Homo sapiens EST sequence
JOURNAL Unpublished (2002)
COMMENT Contact: The Sanger Centre
The Sanger Centre
Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK
Email: humquery@sanger.ac.uk
Sanger Centre name : sccd10818.154136A
Homo sapiens EST sequence. This sequence was generated as part of
The Wellcome Trust Sanger Institute program to identify and
annotate genes in the human genome. Incomplete or unconfirmed genes
are experimentally analysed using a variety of cDNA library
resources. This sequence was obtained from a PCR product generated
from a pool of up to 100,000 cDNA clones derived from
pool_YT_lib_v_SPD cDNA library. Further information can be found at
http://www.sanger.ac.uk/Teams/Team69/.

FEATURES Location/Qualifiers
source 1. .597
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="20"
/clone_lib="pool_YT_lib_v_SPD"

ORIGIN

Query Match 41.0%; Score 322.4; DB 1; Length 597;
Best Local Similarity 99.7%; Pred. No. 4.2e-74;
Matches 323; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy        463 AATGGCTGGCTGTACATCTCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGAC 522
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db         13 AATGGCTGGCTGTACATCTCACCACGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGAC 72

Qy        523 CATTACTCTGAGCTGGCGGATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAG 582
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 CATTACTCTGAGCTGGCGGATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAG 132

Qy        583 AGGGCTGGCCCGCTCCCTGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCA 642
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 AGGGCTGGCCCGCTCCCTGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCA 192

Qy        643 CTCAACTGGAAAGAGCTGGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAG 702
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 CTCAACTGGAAAGAGCTGGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAG 252

Qy        703 TCTCTTCTCAGTGAGGGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAG 762
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        253 TCTCTTCTCAGTGAGGGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAG 312

Qy        763 GCTGTCTCTTTGGATGATGCCTAG 786
              |||||||||||||||||||||||
Db        313 GCTGTCTCTTTGGATGATGCCTAG 336
```

RESULT 14
BU944126
LOCUS BU944126 960 bp mRNA linear EST 18-OCT-2002
DEFINITION AGENCOURT_10545003 NIH_MGC_107 Homo sapiens cDNA clone
IMAGE:6728350 5', mRNA sequence.
ACCESSION BU944126
VERSION BU944126.1 GI:24132945
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 960)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM3049 row: m column: 21
High quality sequence stop: 628.
FEATURES Location/Qualifiers
source 1. .960
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6728350"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_107"
/note="Organ: breast; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN

Query Match 40.8%; Score 321; DB 5; Length 960;
Best Local Similarity 88.4%; Pred. No. 1.1e-73;
Matches 381; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

```
Qy        356 TCATCCGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCC 415
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCATCCGGGAGAGCCAGACCAGGAGAGGCTCTTACTCTCTGTCAGTCCGCCTCAGCCGCC 60

Qy        416 CTGCATCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGT 475
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTGCATCCTGGGACCGGATCAGACACTACAGGATCCACTGCCTTGACAATGGCTGGCTGT 120

Qy        476 ACATCTCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCTGAGC 535
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACATCTCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTACTCT---- 176

Qy        536 TGGCGGATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGCTGGCCCGC 595
                                                  ||||||||||||||||
Db        177 ----------------------------------------------GAGGGCTGGCCCGC 190

Qy        596 TCCCTGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAG 655
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 TCCCTGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAACTGGAAAG 250

Qy        656 AGCTGGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTG 715
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 AGCTGGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCTTCTCAGTG 310

Qy        716 AGGGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGG 775
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 AGGGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGTCTCTTTGG 370

Qy        776 ATGATGCCTAG 786
              ||||||||||||
Db        371 ATGATGCCTAG 381
```

RESULT 15
AL844307
LOCUS AL844307 614 bp mRNA linear EST 30-JUL-2002
DEFINITION AL844307 pool_AK_lib_v_SPD Homo sapiens cDNA, mRNA sequence.
ACCESSION AL844307
VERSION AL844307.1 GI:22019089
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 614)
AUTHORS Ashcroft,K., Bethel,G., Bye,J.M., Howell,G.R., Huckle,E.J. and
Sheridan,E.
TITLE Homo sapiens EST sequence
JOURNAL Unpublished (2002)
COMMENT Contact: The Sanger Centre
The Sanger Centre
Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK
Email: humquery@sanger.ac.uk
Sanger Centre name : sccd10816.154136A
Homo sapiens EST sequence. This sequence was generated as part of
The Wellcome Trust Sanger Institute program to identify and
annotate genes in the human genome. Incomplete or unconfirmed genes
are experimentally analysed using a variety of cDNA library
resources. This sequence was obtained from a PCR product generated
from a pool of up to 100,000 cDNA clones derived from
pool_AK_lib_v_SPD cDNA library. Further information can be found at
http://www.sanger.ac.uk/Teams/Team69/.
FEATURES Location/Qualifiers
source 1. .614
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="20"
/clone_lib="pool_AK_lib_v_SPD"

ORIGIN

Query Match 40.4%; Score 317.4; DB 1; Length 614;
Best Local Similarity 99.7%; Pred. No. 8.8e-73;
Matches 318; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy        468 CTGGCTGTACATCTCACCGCGCCTCACCTTCCCCTCACTCCAGGCCCTGGTGGACCATTA 527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTGGCTGTACATCTCACCGCGCCTCACCTTCCCCTCACTCCAAGCCCTGGTGGACCATTA 60

Qy        528 CTCTGAGCTGGCGGATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGC 587
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTCTGAGCTGGCGGATGACATCTGCTGCCTACTCAAGGAGCCCTGTGTCCTGCAGAGGGC 120

Qy        588 TGGCCCGCTCCCTGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAA 647
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGGCCCGCTCCCTGGCAAGGATATACCCCTACCTGTGACTGTGCAGAGGACACCACTCAA 180

Qy        648 CTGGAAAGAGCTGGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCT 707
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGGAAAGAGCTGGACAGCTCCCTCCTGTTTTCTGAAGCTGCCACAGGGGAGGAGTCTCT 240

Qy        708 TCTCAGTGAGGGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGT 767
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTCAGTGAGGGTCTCCGGGAGTCCCTCAGCTTCTACATCAGCCTGAATGACGAGGCTGT 300

Qy        768 CTCTTTGGATGATGCCTAG 786
              ||||||||||||||||||||
Db        301 CTCTTTGGATGATGCCTAG 319
```

Search completed: November 16, 2004, 23:30:37
Job time : 3256 secs